edited by
Chiara Cirinnà
Maurizio Lunghi

CULTURAL HERITAGE *on line*

Empowering users: an active role for user communities

FIRENZE UNIVERSITY PRESS

CULTURAL HERITAGE ON LINE EMPOWERING USERS: AN ACTIVE ROLE FOR USER COMMUNITIES

Florence, 15th - 16th December 2009

edited by
Chiara Cirinnà and Maurizio Lunghi

Firenze University Press
2010

Cultural Heritage on line : Empowering users: an active role for user communities. Florence, 15th - 16th December 2009 / edited by Chiara Cirinnà and Maurizio Lunghi. – Firenze : Firenze University Press, 2010.
(Proceedings e report ; 69)

http://digital.casalini.it/9788864531878

ISBN 978-88-6453-184-7 (print)
ISBN 978-88-6453-187-8 (online)

Università degli Studi di Firenze
Firenze University Press
Borgo Albizi, 28, 50122 Firenze, Italy
http://www.fupress.com/

Printed in Italy

Promoters of the Conference

Supporters of the Conference

Acknowledgements

Special thanks to the hospitality of Teatro della Pergola.

A special thanks to all the staff working for the organisation of the conference: Emanuele Bellini, Alessio Bertolani, Benedetta Caporioni, Maria Caterina Cunsolo, Verdiana Fontana, Gabrielle Giraudeau, Irene Lisi, Damiana Luzzi, Matteo Paternò, Marco Rufino.

Many thanks to: Firenze University Press, Scuola di Dottorato di Ricerca in "Telematica e Società dell'Informazione" - Università degli Studi di Firenze, Università degli Studi di Roma "Tor Vergata", Fondazione Mediateca Regionale Toscana, Fondazione Giangiacomo Feltrinelli, Sistema Bibliotecario di Ateneo - Università degli Studi di Firenze, Società Dantesca Italiana, Società Internazionale per lo Studio del Medioevo Latino, Fondazione Ezio Franceschini, Memoria Ecclesiae, Federation of Telecommunications Engineers of the Europen Union – FITCE, Associazione Comunicazioni e Tecnologie dell'Informazione – AICT, Master in Management, Marketing e Multimedialità per i Beni e le Attività Culturali - Politecnico di Torino, Fabbrica Europa, Master in Catalogazione Informatica del Patrimonio Culturale - Università Cattolica del Sacro Cuore, sede di Milano, Master in Conservazione, Gestione e Valorizzazione del Patrimonio Industriale - Università degli Studi di Padova, Master in Diagnostica Avanzata per i Beni Culturali - Università degli Studi di Bologna, Master in Indicizzazione di Documenti Cartacei, Multimediali ed Elettronici in Ambiente Digitale - Università degli Studi di Roma Tor Vergata, Master in Restauro Digitale Audio/Video - Università degli Studi di Roma La Sapienza, Master in Studi sul Libro Antico e per la Formazione di Figure di Bibliotecario Manager - Università degli Studi di Siena, Master Internazionale in Information Technology - Scuola Superiore Sant'Anna, Segreteria didattica Facoltà di Conservazione dei Beni Culturali - Università Alma Mater di Bologna, Scuola Normale Superiore, PERCRO - Scuola Superiore Sant'Anna.

Table of contents

Chiara Cirinnà and Maurizio Lunghi (edited by), *Cultural Heritage on line: Empowering users: an active role for user communities. Florence, 15th - 16th December 2009* ISBN 978-88-6453-184-7 (print) ISBN 978-88-6453-187-8 (online)


CULTURAL
HERITAGE
on line
11.00 - 13.30 Plenary Session
13.30 - 15.00 Lunch
15.00 - 18.00 Plenary Session
Back
Exit
15
December


for user communities
INTERNATIONAL CONFERENCE
Florence 15th - 18th December 2009
ssella CAFFO
fano VITALI
nbruno RAVENNI
Exit


Max KAISER
europeana
connect
EuropeanaConnect
Enhancing User Access to European Digital Heritage
Cultural Heritage Online 2009
Florence, 16 December 2009
Max KAISER

The Ente Cassa di Risparmio of Florence represented by myself today is delighted to attend this second edition of the international conference: Cultural Heritage On-line. Empowering users: an active role for user communities, well organized by the Fondazione Rinascimento Digitale and promoted jointly with two of the most important institutions in the survey of the Italian and foreign cultural heritage: the Ministry for Cultural Heritage and Activities, and the Library of Congress.

The conference has involved the participation of several national and international supporters and I would like to thank all of them for the excellent work of collaboration and the extraordinary attendance registered.

The Fondazione Rinascimento Digitale is a Foundation promoted by the Ente Cassa di Risparmio of Florence, that carries out and will carry out more and more a constructive support action to optimize and direct projects concerning the field of new technology for the cultural heritage. We are very glad to see that the Foundation, as years go, has won a high regard and a strong credibility from the most important representatives in the cultural field at the national and international levels.

In conclusion, I would like to thank all the people attending the conference for the great enthusiasm and interest they have showed, and finally I wish to extend my sincere thanks to the Teatro della Pergola for its fine hospitality and helpfulness.

Michele GREMIGNI
President
Ente Cassa di Risparmio di Firenze

Chiara Cirinnà and Maurizio Lunghi (edited by), *Cultural Heritage on line: Empowering users: an active role for user communities. Florence, 15th - 16th December 2009* ISBN 978-88-6453-184-7 (print) ISBN 978-88-6453-187-8 (online)

The second edition of the conference, Cultural Heritage online – Empowering users: an active role for user communities, has been successfully organised in Florence on 15-16 December 2009 by the Fondazione Rinascimento Digitale – FRD in cooperation with the Italian Ministry for Cultural Heritage and Activities (General Direction for Archives and General Direction for Libraries, Cultural Institutes and Authors' Rights), and with the Library of Congress that mobilized all the network of the National Digital Information Infrastructure and Preservation Program - NDIIP partners.
We warmly thank our promoters, the local and regional authorities, and a fantastic group of thirty-five supporters who really were the engine and petrol in spreading information about the event within their user communities, involving cultural heritage institutions and professional associations, technology providers and research centres. A special thanks also to the Director of the Teatro della Pergola for his marvellous hospitality and technical support, and finally to the Palazzo Borghese that was the incredible venue for the gala dinner.
The majority of the success has been possible thanks to all the speakers and chairpersons, as well as to the FRD staff, very limited in number but determined and motivated.

The FRD is a structural foundation promoted by the Ente Cassa di Risparmio di Firenze to support the best adoption of ICT and international standards with a special attention to Internet technologies and digital preservation. The Foundation operates in all of the sectors connected with the production, preservation and diffusion of digital memories by means of research projects and test-beds, management and preservation of digital memories, implementation of innovative applications, and dissemination of this knowledge, with tutorials or on-line training courses.
The FRD invests its annual budget on research projects or prototypes in synergy with the main cultural actors, promotes awareness and training about digital libraries and digital preservation with special attention to young generations.
The main interests and activities of the FRD are in two fields: digital libraries for the cultural and humanities sector, and all the issues related to digital preservation. In the first group, the Foundation has promoted a project to set up a user community about humanities & computing, and has developed an open-source software "Pinakes" to manage digital objects in a very advanced way. Concerning digital preservation, the Foundation is very active and participates in international arenas, for example with the Digital Preservation Europe – DPE coordination action, and with the "Digital Stacks" project, coordinated by the Central National Library of Florence - BNCF, about reliable digital repositories paving the way to the first test-bed for the national legal deposit of digital contents on behalf of the Ministry of Culture in Italy. The Foundation has also promoted an innovative architecture for National Bibliographic Number - NBN persistent identifiers for digital contents, and thanks to the BNCF and the Consiglio Nazionale delle Ricerche – CNR, they have developed jNBN an open-source software already requested by other countries for evaluation. Finally, the Foundation participates with an expert in the PREMIS Editorial Committee managing the development of the standard under the coordination of the Library of Congress. By these three projects, the Foundation attempts to face the challenge of digital preservation with a global and complete strategy.

This year, the conference topics were again related to digital libraries, digital preservation and how Internet is changing scenarios and paradigms, but focussing on the user needs and point of view; so instead of investigating the technology offer, we favoured the works about organisational issues, new Internet schemas and roles, user requirements and

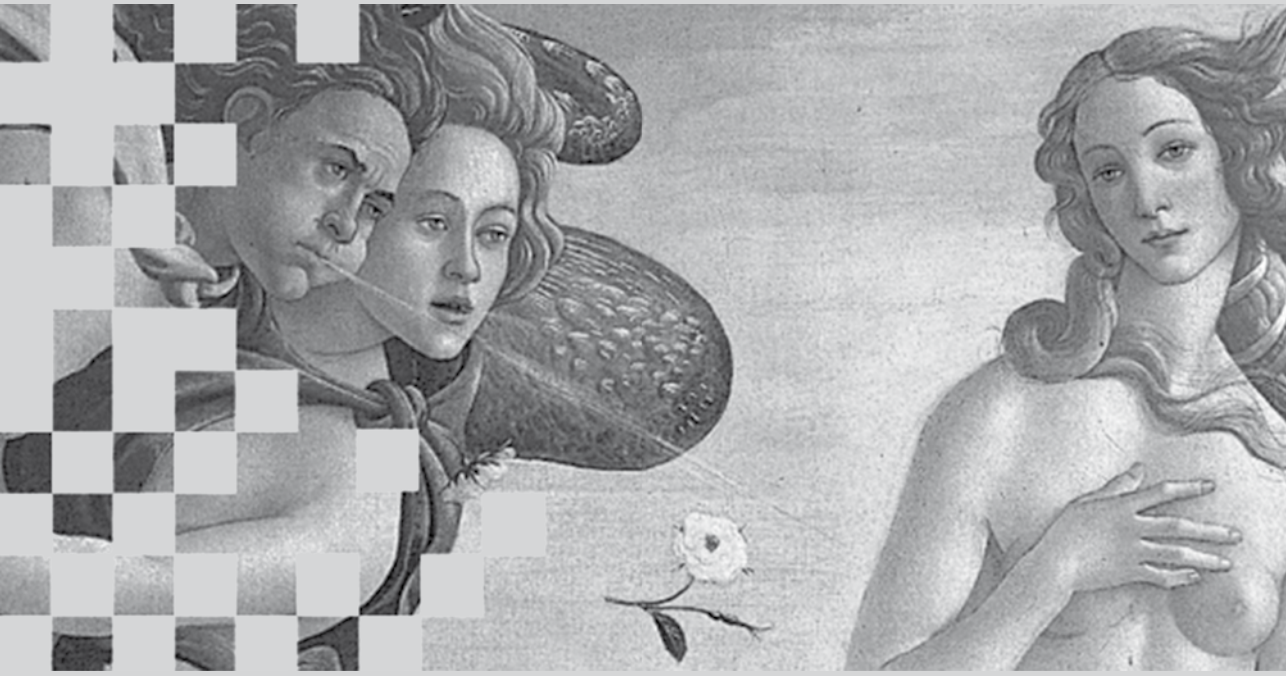
Chiara Cirinnà and Maurizio Lunghi (edited by), *Cultural Heritage on line: Empowering users: an active role for user communities. Florence, 15th - 16th December 2009* ISBN 978-88-6453-184-7 (print) ISBN 978-88-6453-187-8 (online)

constraints, and cultural and economic limits too, that can be seen as obstacles in the way of completely adopting ICT in the cultural heritage sector.
In particular, the first day of the conference proposed eight invited lectures that investigated user needs and expectations, analysing how to better involve users and the cultural heritage community in creating and sharing digital resources. Other topics evaluated were the current trends and use of interactive Web 2.0 tools by cultural institutions, benefits and future opportunities but also limits and constraints for users to create new contents, or possible risks trusting too much any info available on Internet. Some basic concepts underpinning dematerialisation of traditional archives have been presented. The need to choose what digital contents we want to select for long-term preservation, also involves some ethical issues. One important message was that the challenge of the future information society hinges on the use through cooperation among all the sectors, and in particular the current development for Archives, Libraries and Museums - ALM sounds promising.
The plenary session of the second day conference was started with the presentation of national and international scenarios, followed by two thematic sessions with scientific speeches selected through a Call for Papers that observed the advancement of the research on the user-institution relationship towards the development of cooperative Web 2.0 tools and on sustainable digital preservation policies.
The scientific program was enriched by two important training opportunities the day before and the day following the conference: the Tutorial "Long Term Preservation of digital assets: basics, concepts and practices" and the Tutorial"Dublin Core - Building blocks for interoperability".
The conference has seen a great interest not only among the specialists in the cultural heritage field, but there was also a large participation of Italian and foreign students, several European Projects, representatives of local administrations, research centres and the private and corporate sector. There were about 400 people at the conference: one-fourth of those present was formed by students; as for nationality, there were people from all countries: in particular three–quarters of participants were Italians, followed by Americans, Germans, Dutchmen, Estonians, and a varied representation of almost all of the European countries.

In conclusion, many activities have been carried out since the first edition of the Cultural Heritage on line conference in 2006, the Fondazione Rinascimento Digitale has developed contacts and cooperation with cultural institutions worldwide, and we feel confident to be able to work successfully with you all before the next conference in 2012. We'll keep in contact with you before that.

Thanks again to everybody!

Maurizio LUNGHI
Scientific Director
Fondazione Rinascimento Digitale



Plenary Session
Invited lectures
Tuesday 15th December

As we think back to Thomas Jefferson's initial contribution to the Library of Congress, we are reminded that technological progress did not begin with the digital era. But it has been through the continued refinements of digital technologies over the past 20 years that the Library has been most able to make its treasures widely accessible and to share its resources with a national and international audience.
The Library of Congress's embrace of digital technologies is a natural extension of its commitment to innovative access and use of the Library's unique collections. Just as the initial investment in the Library has paid off in myriad ways, so have the ongoing strategic investments in information technology proven essential as a way to bring scholarship to the information leaders of the next century.
As we close the first decade of the 21st Century, the Library is cognizant of its need to continually refine its approaches to engage with an interactive audience fluent with the Web and computing technologies. This refinement is nothing new, as the Library has consistently been a pioneer in identifying these technologies and utilizing them to share its wealth of resources.

1990s Strategy: Increasing Accessibility

The Library of Congress is engaged in a constant process of technological refinement in order to continually maximize the value and accessibility of its collections, but the full embrace of collaborative and interactive digital technologies essentially began in the 1990s.
In 1990, the Library began a pilot called American Memory to digitize some of the Library of Congress's unparalleled collections of historical documents, moving images, sound recordings, and print and photographic media; items that make up the "nation's memory."

Figure 1: Photos from American Memory (http://memory.loc.gov/ammem/index.html)

Chiara Cirinnà and Maurizio Lunghi (edited by), *Cultural Heritage on line: Empowering users: an active role for user communities. Florence, 15th - 16th December 2009* ISBN 978-88-6453-184-7 (print) ISBN 978-88-6453-187-8 (online)

By 1994, as the Internet transformed the presentation and communication of human knowledge, the Library established the National Digital Library Program, a systematic effort to digitize the historical treasures in the Library and make them readily available on the Web to Congress, scholars, educators, students, the general public, and the global Internet community.
The NDLP has been an unqualified success, celebrating its 10 millionth digitized item in American Memory (http://memory.loc.gov/ammem/index.html) in 2004, and continuing to expand online historical content as an integral component of the Library's commitment to harnessing new technology to fulfill its mission "to sustain and preserve a universal collection of knowledge and creativity for future generations."
The iconic materials that comprise American Memory remain highly popular today. Strong interest in the American Memory materials resulted in development of The Learning Page (http://memory.loc.gov/learn/). The Learning Page promotes use of primary sources in the classroom and acts as a teacher's "front door" to Library holdings that enrich teaching experience.
Additionally, the Library worked to make the legislative workings and daily bills of the U.S. Congress more accessible through the establishment of the Thomas website (http://thomas.loc.gov/) in January of 1995.

STRATEGIES FOR THE 21ST CENTURY: ENGAGING THE INTERACTIVE AUDIENCE

The Library entered the 21st century aware of the need to reimagine its traditional roles and responsibilities for new audiences. Building on its earlier digital initiatives, the Library's current digital activities reflect the increasing importance of interactivity in the user experience. These initiatives embrace the core concepts that have come to be known as Web 2.0, with the goal to facilitate interactive information sharing through interoperability and collaboration.
These activities are freeing the Library's wealth of digital content, making it more accessible to a broader community than ever before, both at Loc.gov and in the social networks and Web convergence points such as YouTube, Facebook,

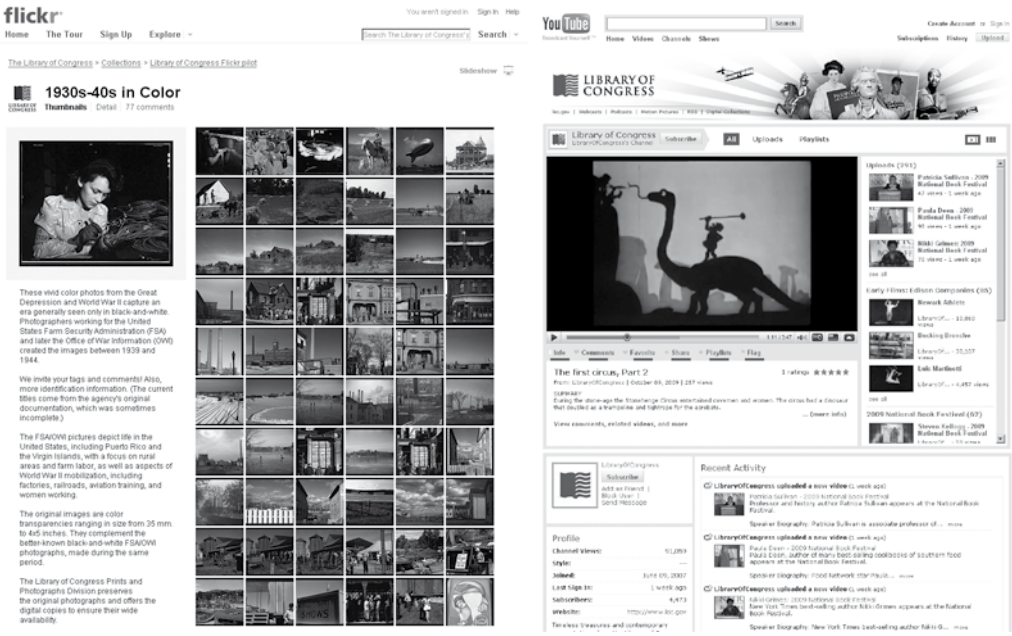


Figure 2: Library of Congress Flickr (http://www.flickr.com/photos/library_of_congress/) and You Tube projects (http://www.youtube.com/user/LibraryOfCongress)

iTunes, Flickr, and Twitter where users now gather. The Library's ultimate goal is to facilitate a deep engagement with its content through actions (such as linking, embedding, tagging or rating) that let users interact with the content in the manner of their choosing.
The Library's Flickr pilot project (http://www.flickr.com/photos/library_of_congress/) provides an illustration of how exposing material through new channels can increase the use of content in new and innovative ways. The project was launched in January 2008 with an invitation to the public to comment on and describe approximately 3,000 historic photos. This opportunity to interact with the Library's collections struck a chord with newly engaged users.
In the first 10 months of the project there were 10.4 million views of the photos on Flickr, resulting in enhancement to more than 1,000 Library of Congress Prints and Photographs Online Catalog records with new information provided by the users. Moreover, the average monthly visits to the PPOC rose 20 percent over the five month period of January-May 2008, compared to the same period in 2007.

The success of the Flickr project inspired the Library to engage in content distribution agreements with Yahoo and YouTube that expose valuable Library content through their channels.
These agreements are models of the way in which government can leverage private sector efforts while ensuring universal access to authoritative content. Any agreement to place content on a non-Library of Congress site must be non-exclusive, and access to the Library's content must be without cost. Notwithstanding advertisements that might appear on various search results pages, an option to control or entirely exclude advertising on the account in close proximity to the Library's content is preferred. The Library also must be clearly identified as the source of the content, and pages with Library content are branded both graphically and through account naming conventions.
The content used in these various pilots is already available on the Library's Web site, ensuring the public can be confident that the material will remain accessible regardless of the long-term viability or future access policies of the non-Library site.
The Library of Congress online video portal on YouTube (http://www.youtube.com/user/LibraryOfCongress), established in April 2009, offers select items from the Library's collections of early motion pictures, along with recordings of Library-sponsored events, lectures and concerts. Through this pilot, the Library - home to over 1.2 million film, television, and video items - is sharing items from its collections with users who enjoy video but may not be aware of the extensive video resources available on the Library's own Web site.
The reach of these new distribution channels is exemplified by the usage statistics of a single video posted to YouTube. The *Rosie the Riveter: Real Women Workers in World War II* video received over 5,000 views on YouTube during April 2009, the first month of the Library of Congress channel launch, and has since garnered over 20,400 views in its first year. It took more than 5 years for the video to reach that level of viewership on LOC.gov. Over 60 percent of the YouTube viewers reached the video through an embedded clip on the Wired.com blog noting the launch of the Library's YouTube channel, demonstrating the potential reach of authorized content sharing.

Network Strategies: Leveraging Experience

In addition to exploring new channels of content distribution, the Library of Congress has engaged with emerging networks of diverse partners. Network strategies allow each participant to leverage the experience, resources and expertise of the entire network community. It has been the Library's experience that networks of collaborating partners increase innovation. The Library has also identified the following benefits to pursuing a network-focused strategy:

- Jointly produced or shared network infrastructure capacity
- Shared product and service innovations and dissemination
- Collective adoption of policy, standards, and best practices
- Increased efficiency in joint problem identification and pursuit of solution paths
- Amassed expert knowledge and knowledge sharing
- Increased efficiencies of pooled capital resources and their distribution
- Increased capacity to target and influence public policy decision makers
- Increased capacity to attract investments in digital preservation

In 2000 the National Digital Information Infrastructure and Preservation Program was created to develop a national strategy to collect, archive and preserve the burgeoning amounts of digital content for current and future generations, a logical outgrowth of the Library's historic mission to "sustain and preserve a universal collection of knowledge and creativity for future generations."
NDIIPP has deeply engaged the Library in collaborative partnerships, with the understanding that digital stewardship on a national scale depends on public and private communities working together. To that effect, the Library has built an ongoing preservation network of over 150 partners from across the nation to tackle the challenge of preserving and providing enhanced access to digital content of importance to the nation.
This approach encompasses a new catalytic role for the Library, with NDIIPP's strategies for success exemplifying the Library's new roles within developing networks of likeminded partners:

- Take early action
- Learn by doing
- Be catalytic
- Act collaboratively
- Engage Multiple platforms
- Leverage Natural, Distributed Networks
- Share Resources

No single organization has the resources and the ability to preserve all the data needed by policy-makers, planners, and scholars. NDIIPP is leveraging the expertise and existing work being accomplished by federal agencies, universities, the private sector, and not-for-profit organizations. By working together, the network can identify and maintain the content, including geospatial data, which will allow complex, temporal analysis of historical data in the future.
Another model of networked collaboration in which the Library of Congress participates is the International Internet Preservation Consortium (IIPC). The IIPC enables the collection of a rich body of Internet content from around the world to be preserved in a way that it can be archived, secured and accessed over time. The group fosters the development and use of common tools, techniques and standards that enable the creation of international archives. In addition to the collaborative work within the consortium, the IIPC also encourages and supports national libraries everywhere to address Internet archiving and preservation.
The twin goals of expanding online offerings through new channels while also embracing networks of diverse partners come together in the World Digital Library, a project being developed with the cooperation of the United Nations Educational, Scientific and Cultural Organization and 32 other partner institutions.
With 55 partners in 35 countries as of October 2009, the WDL makes available on the Internet, free of charge and in multilingual format, rare and unique primary source materials that reflect the history and cultural heritage of the participating countries. The WDL promotes international and inter-cultural understanding and awareness, increases the diversity and volume of cultural content on the Internet, provides resources for educators, scholars and general audiences, and builds capacity in partner institutions in ways that will narrow the digital divide within and between countries.

Figure 3: The World Digital Library (http://www.wdl.org/en/)

Collaboration and Interaction: Strategies for the Future

These recent collaborative initiatives have increased the volume of content, the variety of formats, the kinds of user communities, the personalization of the content and the ability of the Library of Congress to influence and support scholarship throughout the world.

Future initiatives will explore the how the Library acquires and preserves content that exists in disparate locations around the world, brought together through new collaborative methods of creation. The "crowdsourcing" of news reporting is one example of how the traditional channels of communication are changing as digital technologies enhance user's ability to both create and consume content. The Library will also be called upon to support new, collaborative methods for learning, education and research.

The technological environment is rapidly changing, and the Library is changing with it, continuing to develop new strategies to showcase its collections based on leveraging opportunities for collaboration and interaction where they exist while grappling with the exponential growth in content and the concomitant increasing user engagement.

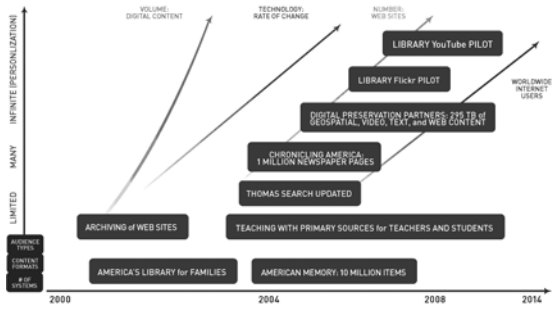


Figure 4: The Library of Congress's Digital Content and Interactivity

Digital preservation can be defined as the whole of the principles, policies, rules and strategies aimed at prolonging the existence of a digital object by maintaining it in a condition suitable for use, either in its original format or in a more persistent format, while protecting the object's identity and integrity, that is, its authenticity.[1] One might ask why authenticity is an issue for all kinds of digital objects. While it is obvious that authenticity is a necessary requirement in the preservation of records, the value of which as sources of evidence resides in their trustworthiness, why would it be for other types of digital heritage, such as publications, works of art, or games?

Although a digital entity that does not qualify as a record is not conceptually linked to the idea of trust, it still needs to have an identity that is certain and indisputable, and its manifestation must be stable, always equal to itself, intact. And here lies the problem, because, traditionally, these qualities reside in the original, the perfect entity, the first complete item to be issued, released or made public (be it a unique one or one of many), or in an authentic copy of the original generated by a person with the authority to do so. However, as the concept of original disappears in the digital world, where can we look for the assurance that what we observe is what it claims to be?

We have no other choice than make inferences based on a variety of circumstances, but primarily on the integrity of the environment in which the digital entities in question reside and of the processes aimed to maintain them and to ensure the accountability of the person or organization responsible for them. This means that institutions must create mechanisms that allow for the determination of authenticity based on the trustworthiness of the source of the digital entities and the chosen method of their transmission through time, and then adopt the necessary strategies to preserve them in a sustainable way.

The selected preservation methodology must allow the preserved entities to continue to be readable and useable regardless of any technological changes to the underlying hardware or software environments, and the preserving organization to account for these changes, so that the entities may continue to be migrated to newer platforms as needed to avoid technological obsolescence.

To illustrate the issues digital repositories of cultural heritage will have to deal with, I am going to discuss an InterPARES 3[2] Case Study called "cIRcle." cIRcle is a digital repository for the management, dissemination and preservation of the intellectual output of a university and its community members.

A university institutional repository is defined as "a set of services that a university offers to the members of its community for the management and dissemination of digital materials created by the institution and its members."[3] Although their creation has been predicated in large part on the requirement of making available to the public research products that have been developed with the support of public money, and on the benefit to the university of showcasing its research, institutional repositories (hereinafter IR) have emerged in North America and Western Europe primarily because they are regarded by the university communities as a means of having access to products of scholarship and research, and as a locus for preserving such products and maintaining access to them over the long term.[4] Therefore,

[1] See the InterPARES Project Terminology Database at http://www.interpares.org/ip2/ip2_terminology_db.cfm.

[2] InterPARES 3 is the third phase of the InterPARES Research Project (1999-2012). While the first two aimed to develop theory and methods of authentic digital long term preservation, the third is testing the findings of the previous two in real situations. See http://www.interpares.org/ip3/ip3_index.cfm.

[3] Clifford A. Lynch, "Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age," ARL Bimonthly Report 226 (2003): 1-7, available at http://www.arl.org/bin~doc/br226.pdf.

[4] See Clifford A. Lynch and Joan K. Lippincott, "Institutional Repository Deployment in the Unites States as of Early 2005," D-Lib Magazine 11 (September 2005), available at http://www.dlib.org/dlib/september05/lynch/09lynch.html; and Gerard van Westrienen and Clifford A. Lynch, "Academic Institutional Repositories: Deployment Status in 13 Nations as of Mid 2005," D-Lib Magazine 11

Chiara Cirinnà and Maurizio Lunghi (edited by), *Cultural Heritage on line: Empowering users: an active role for user communities. Florence, 15th - 16th December 2009* ISBN 978-88-6453-184-7 (print) ISBN 978-88-6453-187-8 (online)

in the past few years, the IRs have accumulated not only preprints and post-prints of articles, books, theses and dissertations, but also raw data files resulting from research, working papers, course syllabi, class notes, handouts, students' papers, committee meetings agendas and minutes, unpublished conference presentations and several other types of documentation that fall under the category of personal and university *records,* not only *publications*, and are preserved by the university archives and special collections. This mix of documentation and data creates severe challenges to the IR's continuing access and preservation (the very reasons why they exist) from several points of view.

The InterPARES research project has demonstrated that it is not possible to preserve digital materials, but only the ability to reproduce them. Reproduction involves different activities at different times in the life of the material. In the initial few years, it may consist simply of retrieving and reassembling the digital components that constitute the object to generate a copy or, if the object is technologically complex, as in the case of interactive and/or dynamic documents from the visual and performing arts and from the sciences, it may require its *re-creation*. However, when the digital format becomes obsolescent, it is necessary to either migrate the digital object to a newer technology by changing its architectural structure or, in some cases, to emulate the behaviour of the old technology to access the object. Regardless, throughout the existence of the object, ongoing copying and transformative migration[5] are required for reasons of security (which is based on redundancy) and of continuing access. These activities raise several issues, among which the paramount ones are those of authenticity and intellectual rights. The authenticity of digital material is dependent upon the maintenance through time of its identity and of its integrity[6]. The intellectual rights of the copyright owner are attached to the authentic version of the digital object and, specifically, to its documentary form, which is defined as the rules of representation that govern the expression of the ideas of the author in a stable and fixed manner.[7]

Intellectual rights comprise several types of rights, but among them the ones that are affected by long-term preservation by means of constant transformative migration or emulation are the two major groups of intellectual rights: economic rights and moral rights. Economic rights are those that enable the copyright owner (not necessarily the author or creator) of a work to make commercial gain from the exploitation of that work.[8] Moral rights are those rights that the author or creator retains (regardless of whether the author still retains the economic rights) over the integrity of a work (rights of reputation) - such that no one, even the copyright owner, is allowed to distort, mutilate or otherwise modify the work in a way that is prejudicial to the author's honour or reputation; the right to be associated with the work as its author by name or under a pseudonym and the right to remain anonymous (rights of attribution); and the right to refuse to allow the work to be used in association with a product, service, cause or institution in a way that is prejudicial to the author's honour or reputation (rights of association).[9]

A recent census of college and university IRs in the United States has found that 70.8% of them do not have a policy for licensing content. In addition, there is no mention in the literature concerning IRs of the issue of authenticity through time, and none of them appears to have strategies in place for long-term preservation.[10] This is probably due to the fact that it is the belief of those who manage an IR that its content exists somewhere else, which is a safe presumption for

(September 2005), available at http://www.dlib.org/dlib/september05/westrienen/09westrienen.html.

[5] Transformative migration is defined as "The process of converting or upgrading digital objects or systems to a newer generation of hardware and/or software computer technology" (InterPARES 2 Project, "Terminology Database: Glossary," available at http://www.interpares.org/ip2/ip2_terminology_db.cfm). The effects of transformative migration of the digital materials in an institutional repository are an important consideration insofar as any new additions or modifications to an existing work (even a work already in the public domain) can trigger new copyright considerations.

[6] Luciana Duranti (ed.), The Long-term Preservation of Authentic Electronic Records: Findings of the InterPARES Project (San Miniato, Italy: Archilab, 2005).

[7] Luciana Duranti and Kenneth Thibodeau, "The Concept of Record in Interactive, Experiential and Dynamic Environments: the View of InterPARES," Archival Science 1 (2006): 13-68.

[8] Michael O'Hare, "Copyright and the Protection of Economic Rights," Journal of Cultural Economics 6(1) (1982): 33-48.

[9] Mira Sundara Rajan, "Moral Rights in Information Technology: A New Kind of 'Personal Right'?" International Journal of Law & Information Technology 12(1) (2004): 32-54.

[10] Elizabeth Yakel, Soo Young Rieh, Beth St. Jean, Karen Markey and Jihyun Kim, "Institutional Repositories and the Institutional Repository: College and University Archives and Special Collections in an Era of Change," The American Archivist 71(2) (2008): 344.

preprints and post-prints, but certainly not for all those digital objects that are unique and often qualify as records, such as the official copy of theses, professors' and students' papers, etc.
Given the situation described above, it is necessary:
to identify in which way digital preservation strategies as recommended by the major international research projects on the subject may infringe existing intellectual rights (economic and moral) legislation as it applies to published and unpublished material;
to establish what long-term preservation measures would be possible in the context of the existing legislation and to test them on IRs in course of development to assess their impact on the continuing authenticity and accessibility of the digital material; and to determine what changes to the law are required to ensure that the proper long-term digital preservation strategies can be applied so that the research output of universities can remain attributable and accessible in its authentic form for as long as needed.
In order to do so, the InterPARES 3 project has selected as a case study an institutional repository called cIRcle, at the University of British Columbia (UBC)[11]. As stated in the brochure publicizing it to the UBC faculty and students, cIRcle assembles various communities and collections. Communities are UBC departments, labs, research centres, schools and other administrative units. Within cIRcle, each community oversees one or more of its own collections, which contain items submitted to the IR. As currently envisioned, cIRcle's operational goal is to be able to accept, preserve indefinitely and provide continued readability and accessibility to virtually all published and unpublished digital objects created in any file format by or on behalf of the University, its faculty, staff or students - including preprints and post-prints of academic journal articles, other items such as theses, dissertations, departmental publications, technical reports, bulletins, conference proceedings, course notes and other learning objects, and raw research data. cIRcle has yet to develop, articulate and implement a maintenance plan that addresses this ambitious goal, and it has not attempted yet to address the issue of the protection of economic and moral rights in the context of long-term preservation. This situation makes of it the ideal candidate for the development of a preservation strategy for IRs that is sensitive to intellectual rights issues and for the testing of such a strategy.
As of November 6, 2009 there were 14,073 items in cIRcle totaling 130GB. This material is stored in DSpace, which is a database with a set of services to capture, store, index, maintain and make accessible a variety of entities in a digital format over the internet utilizing a controlled set of workflows and access permissions. Dspace is an open-source application, freely accessible at sourceforge.net, one of the largest open source software repositories on the net. It is written in Java, providing broad based support and compatibility with a broad base of internet browser; for a database back end it uses either Oracle – the industry leader – or Postgres, an open source relational database. The fact that DSpace is an open source application is good because of the authenticity issue.
It was stated earlier that the preservation of authenticity, to which all intellectual rights are linked, requires the protection of the identity and integrity of the material. Identity is not difficult to maintain overtime if the appropriate set of metadata is attached to the various digital entities and kept inextricably linked to them. Integrity is problematic not just to protect, but also to prove, because one has to rely on the integrity of the environment in which the entities reside. It is very hard to assess the integrity of an environment that is proprietary. In contrast, open source satisfies the legal requirements of objectivity, transparency, verifiability and repeatability for any process that is carried out in a digital environment[12].
In DSpace, the records themselves are embedded in a hierarchical folder structure based on the collection. Contained within each folder is the original bitstream, a full text extract of the contents (used for searching), a thumbnail of the record for web presentation, and copyright information on the record. DSpace uses Preservation Services modules to verify the integrity of the stored files (utilizing a checksum to look for file corruption or alteration) and media filters to

[11] See https://circle.ubc.ca/.
[12] See for example Brian Carrier, "Open Source Digital Forensics Tool. The Legal Argument," available at http://www.digital-evidence.org/papers/opensrc_legal.pdf, p. 7; and Erin Kenneally, "Gatekeeping Out Of The Box: Open Source Software As A Mechanism To Assess Reliability For Digital Evidence," Virginia Journal of Law and Technology, 6 Va. J.L. & Tech. 13 (2001). Available at http://www.vjolt.net/vol6/issue3/v6i3-a13-Kenneally.html, para. 34-35. See also Andrea Ghirardini and Gabriele Faggioli, Computer Forensics (Milano: Apogeo, 2007), p. 230.

define file conversations. Records are accessed through the web via a persistent web address that allows researchers to link directly rather than having to use a database search every time. The bitstream format contains information on how the material in that format is to be interpreted, allowing for control and granularity. For example, .doc may refer to more than one version of word, each of which presumably has different characteristics and functionality. Each bitstream format also has support level associated with it indicating how likely it is that the format will be accessible into the future given the toolset currently available to the system[13].

Digital forensics experts value open source, which, at the same time, allows modification and encourages dissemination, thereby making it possible to submit the software together with the digital entities presented as evidence, so that their accuracy can be tested promptly by anyone at any time. This is especially true when conversion or migration occurs, because it would allow a practical demonstration that the software could not simultaneously manipulate the content of the files while copying them and that nothing could be altered, lost, planted, or destroyed. Finally, open source is preferred because of the possibility of exchange of evidentiary material between the parties in the course of e-discovery.

Why should we care about issues of evidence and discovery? Because it is more than certain that, if an author feels that his or her intellectual rights have been infringed by the preservation measures taken by the institutional repository, he or she will want to see the issue solved in court. Undoubtedly, cIRcle procedures are in conflict with copyright legislation also regardless of preservation methods. This is because one has to consider that acting within copyright is different from policing copyright. Items are generally posted to cIRcle by the author of a digital entity or the author's representative. Each submission requires the depositor to authenticate his/her authority to submit this work. cIRcle staff don't have enough time to verify copyright ownership for each item submitted, so cIRcle has to rely on the declaration of the depositor in order to operate. This is an act of faith, but it is a necessary one. Provided that cIRcle removes work upon notice of an infringement, and provided that cIRcle did not publish the infringing work knowingly, cIRcle should be protected from prosecution. Another issue has to do with materials that are scanned and uploaded to an IR. cIRcle's retrospective theses project involves the deposit of digitized theses and dissertations originally archived in print. It can be difficult to contact the authors of these items to obtain their permission to deposit. When authors cannot be contacted after due diligence in attempting to notify them, cIRcle may choose to proceed with publishing their work online. They do so assuming implicit permission by dint of the university's prevailing stewardship and provision of access to the item. Should the author request removal of the work from cIRcle after it has been published online, cIRcle remains obliged to honour the author's wishes and remove the work from its holdings.

But these are minor issues with respect to the problems presented for intellectual rights protection by the preservation strategies that cIRcle will have to adopt in the presence of a legislation that is still much behind the development of technology. cIRcle will have to begin separating the protection of the moral rights and that of the economic rights. To do so, its strategy will need to distinguish data integrity, which means that the content of the entities in the repository is not modified accidentally or intentionally during the regular maintenance and use activities, from duplication integrity, which means that the process of creating a duplicate of the data for preservation does not modify either intentionally or accidentally the form and composition of the original entity. This reproduction would either be based on the principle of non-interference, which involves a non-transformative conversion, or on the principle of identifiable interference, which means that the method used does alter the entities but the changes are identifiable[14]. The application of both principles, by ensuring the creation of authentic copies, would allow for the protection of moral rights, which cannot be renounced. As it regards economic rights, any preservation activity would infringe the law as it stands now, thus, the only solution at this time is to obtain the permission of the copyright owners.

The InterPARES Project has made a submission to the Commission of Industry Canada and the Department of Ca-

[13] The three levels are Supported, Known and Unsupported. Supported are those file types that the institution has reasonable level of confidence that they have the tools and/or techniques available to progress the files through future technology changes. Known formats are those that are recognized by the institution and attempts are being made to create or obtain the tools necessary for future migration/access. Unknown file formats are those that will be preserved at the bitstream level only; it will be up to the researcher to obtain the software/tools necessary to view the files.

[14] E. Casey, "Error, uncertainty and loss in digital evidence," International Journal of Digital Evidence 1. 2 (2002).

nadian Heritage responsible for updating the Canadian copyright act, requesting that specific attention be given to the problems presented by the long term preservation of authentic digital entities.[15] In the meanwhile, it is conducting an inventory of all the items in cIRcle to identify their nature and characteristics, content, current licence, attached digital rights management, etc. in order to develop an intellectual property policy and a preservation plan consistent with it. The research conducted on cIRcle and its results will be accessible on a dedicate web site named "University Institutional Repositories: Copyright and Long Term Preservation," accessible at http://uir-preservation.org/[16].

[15] Copyright Consultation statement available at: http://copyright.econsultation.ca

[16] For more information on copyright and moral rights as related to preservations see: Austin, G. W., "The Berne Convention as a Canon of Construction: Moral Rights after Dastar"NYU Annual Survey of American Law 61 (2005): 101-139; Burkitt, D., "Copyrighting Culture – The History and Cultural Specificity of the Western Model of
Copyright," Intellectual Property Quarterly 2 (2001): 146-186; Christie, A., "Reconceptualising Copyright in the Digital Era," European Intellectual Property Review 17(11) (1995): 522-530; Davies, G., Copyright and the Public Interest, 2d ed. (London: Sweet and Maxwell, 2002); Dreier, T. K., "Adjustment of Copyright Law to the Requirements of the Information Society," IIC 29(6) (1998): 623-639; Dreier, T. K., "Authorship and New Technologies from the Viewpoint of Civil Law Traditions," IIC 26(6) (1995): 989-999; Frow, J., "Public Domain and Collective Rights in Culture," Intellectual Property Journal 13(1) (1998): 39-52; Gervais, D., "Canadian Copyright Law Post-CCH," Intellectual Property Journal 18 (2004): 131-167; May, C., "Why IPRs are a Global Political Issue," European Intellectual Property Review 25(1) (2003): 1-6; New York Times Co. v. Tasini, 533 U.S. 483 (2001).14; Padfield, Tim, Copyright for Archivists and Users of Archives (London: Facet Publishing, 2004).
Rigamonti, C. P., "The Conceptual Transformation of Moral Rights," American Journal of Comparative Law 55(1) (2007): 67-122; Robertson v. Thomson Corp, 2006 SCC 43.15; Sterling, J. A. L., "Philosophical and Legal Challenges in the Context of Copyright and Digital Technology," IIC 31(5) (2000): 508-525; Vaver, D., Copyright Law, Essentials of Canadian Law Series (Toronto: Irwin Law, 2000).

It has been widely assumed that digitisation of analogue contents is the most effective process for preserving contents from their natural and inevitable physical decay. From a technical point of view, and independently of the risks incurred within the digital preservation process, this assumption is true since digital replication permits to reproduce contents without altering them. However, from a more strict ethical approach on the nature of the content to preserve, it is important to question the continuity of the original content through its technological transformations and to analyse the concept of content equivalence.

Within the domain of text, the possibility to separate the information from the physical object containing it has been part of a long known strategy to reproduce written material without any major losses (except copying errors). With analogue media and images (as well as sounds), the fact that the content is imbedded in the carrier conditions the quality of the information and creates new replication paradigms, difficult to analyse and quantify efficiently. Independently of procedure standards, which may guarantee the quality of the process, it becomes extremely difficult to evaluate what quality means, mainly for images and sounds whose initial quality may be considered as very poor regarding today's technology. The problem goes further beyond the digitisation process, when we encounter contents, which have gone through a time decay process that has deteriorated the original state of the content. A major question arises then: which was the original state of a decayed content? Should we try to recover the initial state or even improve it?

The discussion happens also in the Digital world, where migration processes bring small deformations which may be important through time; where transcoding is a necessary action to maintain contents accessible, however it may introduce artefacts that can be annoying with time.

Content representation

Content representation is associated to the concepts of original and copy; the original is the work of a content creator (artist, author, or whatever name is applied) who is the originator of the work and claims its ownership (several authors may also be concerned plus other right-holders). A copy is a reproduction of the original, which conserves some or all of the features of the original content or object. A representation is a more or less accurate description of something; representations are used as equivalents of the original object even if the representation is very poor (a cinemascope film viewed on a small television set with transmission problems is still considered as viewing the film). Representations are very useful objects to help us deal with contents or to simplify the access to them.

The difficulty is that: in some cases the distance between the original and its representation may be clear for our perception (it is evident for all that it is not the same thing to go to the Louvre to see the Mona Lisa by Leonardo, than to see the image on a cookie package); however in other situations the difference is not appealing for our perception: a photocopy of a book is considered as being the same thing as the book, a highly quality degraded photography is considered as having the same value as the original photography itself.

Depending on the domain originals or copies, or representations; don't have the same implication and value: if we take four types of cultural domains as Museums, Libraries, Archives and Audiovisual collections, the concept of original or copy has different implications:

- Museums tend to contain physical objects; each object has an author or several authors, which can be known or unknown. Often high value is associated to the objects, which are mainly visible in the Museum's site. In this context copies are representations of the objects that serve to inform the visitors about the object; however viewing the physical object is the main objective for the Museum.

  Copies are considered reductions of the original and our perception identifies them as representations of it. There are also physical copies of objects, and they can be very tricky to identify, however this is outside of the scope of this analysis.

Chiara Cirinnà and Maurizio Lunghi (edited by), *Cultural Heritage on line: Empowering users: an active role for user communities. Florence, 15th - 16th December 2009* ISBN 978-88-6453-184-7 (print) ISBN 978-88-6453-187-8 (online)

- Libraries tend to contain written material (books or manuscripts), the books are the carrier for information; however the information can exist independently of the book and be transmitted through copies. Since writing is a coding system for information, the copy is as good as the original (mistakes may be introduced in the process)
  Publishing is based on the principle of making multiple copies of the same information providing the original information is not modified, whatever way the information is presented (or reformatted in shape, typography or font size)
- Archives contain collections of objects, which may include physical objects, written material, audiovisual elements, etc. The main characteristic is that objects are related together through a collection, which may contain many authors and an owner of the collection. The links among the objects are as important as the objects themselves and the concept of original is related to these links that give coherence to the collection. Within the collection each individual object has the same copying issues as those seen in other domains.
  Parts of the archival material may be copied, however the main concepts are the integrity of the whole collection and the links among the objects. These links are a knowledge that is added to the material; copies can be made of the structure as well as the objects.
- Audiovisual collections contain media (film, video, audio) in which information is linked with the media ("the media is the message" as Marshall McLuhan said). The notion of author is highly variable (several right owners generally). The main characteristic is that the deterioration of the media will have as a consequence the deterioration of the information or content. Copying an original normally gives a highly similar equivalent of the original (with loss in the analogue world and theoretically no loss in the digital world).
  A copy is a new entity of the work; however there may be slow and imperceptible changes and degradations. A strongly degraded copy is still considered as a representation of the original.

Living with physical decay

Time is a perpetual threat for any kind of object: books, archives, media, physical objects decay slowly or faster and that is the fate of our physical world. Probably the most efficient way to conserve information is to regularly copy it on a new carrier until it starts decaying and the process of copying is started again. We have no original books from antiquity, however the writings of Plato or Aristotle are still available through the slow and tedious process of successive handmade copies until the arrival of print. These copies probably introduced small or important errors, to which the process of translation should be added, enlarging the divergences with the original source. We can still access those writers and understand their words and ideas, and sometimes track the different versions and transcriptions of the work. This process, when applied to digital contents and even to analogue to digital transfer, is called migration and it is largely accepted that it is the best solution in order to keep digital contents accessible. As with book copying, it is important to keep track of all the migration actions in order to evaluate eventual deformations or degradations. Concerning physical objects, specific actions have to be undertaken to minimize the action of time regarding their integrity. Restoration has to be done when parts of it fall apart and specific temperature and humidity as well as light environments have are foreseen in order to reduce the action of elements on them.

Preservation actions have to be organised and planned because protecting objects under special conditions, doesn't stop the action of time and, concerning analogue media for example, there is no way of stopping the decay process. The digital domain brought the hope for a new world in which there would no more be information transmission problems since copies were exact replicas of the original content. Just by regularly replicating would bring eternal life to digital contents. Reality shows that the replication process is perfect in theory but in practice many small distortions may appear. This has brought the need for new words to characterise needed actions: authenticity, consistency, replication, integrity…

A second major problem is the format in which the content is codified, which should be able to be accessed in the future. If no specific transfer actions are undertaken in order to maintain the feedback environment needed in order to access information; content may be well preserved however no access will be possible since the "decoder" that permits information to be interpreted by our senses, no more exists. Some words are associated to this situation and actions, as: transcoding, obsolescence, emulation, virtual universal formats…

To summarize the problems, digital copying gives identical replicas of documents; however it may introduce small differences and the document may not be accessible in the future because of compatibility aspects. Preservation as a paradigm has changed in the digital world: keeping a digital object in a secured and controlled physical environment,

doesn't guarantee by any means that the document will be accessible in the future. Preservation ceases being "passive", to become "active", which means continuously checking the document and migrating it to a new environment or format when necessary. The major problem addressed is longevity and the capacity to organise preservation in such a way that any document may be accessible in any future.

How to ethics intervene?

Ethics study which properties, if any, are responsible for the truth or validity of something. This kind of loose definition is closely related to the concept of author, and the moral rights that being an author confers: "you should not modify my work" is one of the strong aspects of authorship and has implications in the use and diffusion of contents (unless it is put on Creative Commons with that property!). Under this condition, some questions should be addressed concerning analogue to digital migration, as well as digital transcoding or restoration. Within the Audiovisual domain (which includes cinema), where information and carrier are closely interrelated and dependent, these issues may have strong implications.

Going from analogue to digital represents a huge paradigmatic change in the way information is coded. Even if a film or video and its digital equivalent look very similar, they are completely different from a technical point of view. Our perception considers both as being equivalent, which is the starting point for an analogue to digital high-quality migration. Almost always, information is compressed in order to optimise storage space and transmission bandwidth; this process is applied as long as it doesn't affect the visual rendering of the original image. Even if the image is not compressed and the digital definition is as high as the analogue definition, the technical principles underlying image representation are different. Curiously the "authenticity" of the digital replica is not questioned and authors have easily accepted migration, considering it as an added value in terms of longevity (the process is accepted with no restrictions in the video and sound domain, however within the film domain only nowadays it is considered that the digital replica of a high quality cinema film is equivalent or at least acceptable when comparing it with the original 35 or 70mm film).

Preservation exists since the beginning of film, radio and television. It happened accidentally, when carriers were kept somewhere and then rediscovered and reused; or voluntarily, when content were organised in collections associated with documentation and conserved in specific environments. In any case, these contents have undergone the challenge of time and our knowledge about preservation comes mainly through the experience accumulated while trying to keep them accessible. Since the beginning of audiovisual preservation, the quality of images and sounds has continuously improved. At the same time, our senses have become adapted to high quality audiovisual information and, even if images of the past are associated with archives and that gives them an emotional value, their quality is very poor from a technical point of view. However a question always remains: am I viewing the image as it originally was, or am I viewing a degraded image of the original image? This question has no clear answer, since in some cases the action of time is perceptible in what is called image defects (scratches, drop-outs, dirty splits), in other (many) cases, a slow and progressive degradation of the image is difficult to detect and there is no way of knowing what the original quality was. This has implications in the "restoration" domain, where the ambition is to restore the initial quality of an image or a sound. Defects are corrected in order to achieve what is considered as being the original quality. This process can go beyond the hypothetical "initial quality" and the image can be improved thus making it look better than it originally was. The quest for the lost quality has gone beyond the initial ambition; in the age of High-Definition, of 3D images, even quite recent images can seem of very poor quality in comparison with today's productions. There is then a tendency to improve the original quality or to upgrade it in order to be acceptable when used in new contexts.

When images or sounds have a high an "artistic" value (as in cinema and music) it is considered that restoration should not alter the original quality of the content (including its initial technical defects). When addressing other kind of contents, where the notion of author is less present (television or radio programs, recordings of famous voices or moments) there is a tendency to try to improve the quality to a higher point than it initially was. In other words aging brings extra value to films and music, and implies bad quality for other domains.

Two different conceptions

Even if law establishes quite clearly what author rights are and the conditions under which contents should be used; there are in fact two kinds of ethics: institutional ethics, where institutions have precise missions regarding content

preservation and access with a very high respect to author's rights and the context in which contents are used. On the other hand there is what users daily do with contents; they circulate with little control and concern about author rights and it is considered that contents should be free for all within the web space. Ethics here are completely different!
Authors are more concerned with the circulation of their contents and the respect of their rights than with the quality of what circulates or with the ethic implications of going from analogue to digital. A certain way of thinking has prevailed in the last years concerning quality: it is considered that bad quality copies protect contents, since the originals are of a much higher quality, meaning that the original is safe somewhere and if bad copies circulate it has lesser implications. This doesn't prevent users from downloading bad quality copies and to be rather satisfied of viewing a very bad quality copy of a new film instead of going to the cinema. Bad qualities were the consequence of narrow bandwidths, but as the quality of transmission improves, so does the quality of the images that circulate, so the bad quality protection will be without much sense in the future.
As quality advances authors and content holders may be more concerned about quality and the thin difference that exists in many domains between originals and copies will progressively fade out. The concept of author and the concept of work are also evolving and will need in the future adaptations to continuously changing technical, legal and user environments. Systems for controlling the circulation of contents on the web are quite widely used today, however the main concern could be in the future to be capable of identifying what is circulating and thus finding efficient ways to remunerate authors instead of trying to block the circulation of protected contents.
The role of institutions in this changing environment is difficult to anticipate. They have a role of keepers of cultural contents and they need to develop strong preservation strategies in order to guarantee that the original contents or certified copies are preserved forever. In parallel, they have to interact and exchange with the external world while being the warrants of authenticity and of respect to authors and other right holders.

The world is undergoing several radical changes.

The effects of the Financial Crisis scourging the economy are amplifying the magnitude of phenomena - dating back to early 90s – bound to the evolution of digital technologies and to the widespread of the Internet, which have had a profound effect on people behaviour and on social dynamics.

Said trends impacted strongly global markets, either by changing the competitive context or by opening up local markets to a worldwide dimension, introducing new challenges both for goods producers and for service providers: companies have to deal with global audiences and worldwide competitors never seen before, restructuring their offering to meet the ever-changing needs of their customers.

Goods producers are facing a massive introduction of new technological solutions at a pace which makes hard to exploit them adequately. Service providers are confused by a smart, active audience with changing tastes and needs; more specifically, content providers and media companies are involved in a deep crisis tied to the failure of copyright-bound business models in favour of an economy based on almost-free and/or user generated contents.

This change in competitive context, in short, has opened up a panorama both of big opportunities and of tremendous threats.

From the point of view of economic operators, especially the content providers and those active in the so-called *knowledge economy*, the challenge is to learn how to cope with the high rate of change and heterogeneity of customers, namely to produce content that is capable to engage them, becoming memorable and useful.

Besides mere technical and engineering issues - not pertaining the object of this document - there's the need to find clues of alternative ways to relate to customers and markets, both in the analysis of problems and requirements, and in the ways new technologies are used and perceived.

It's my opinion, in fact, that – as of today - users (and customers) play a role in innovation processes that is much important than ever, because of their new source function.

For explanation's sake, we introduce here 4 kinds of innovation:

- Product Innovation;
- Process Innovation;
- Organisational Innovation;
- "Customer-enabled" innovation.

Product, Process and Organisational innovation are the traditional processes involving technicalities and internal aspects of companies management; "Customer-enabled" innovation, vice-versa, leverages the fertility of the "common wisdom" in order to find clues of flaws or improvements that can be made to a product or service.

This can be achieved mainly in two ways, both improved by the use of bidirectional mass communication channels:

- analyzing the way customers use and perceive a given technology or solution, as a source of invaluable insight for the inception of new products, or for the design of successive iteration of the same product or service – which we call "Use Innovation";
- analyzing the way groups of individuals react to complex problems in order to discover smart solutions, to build further improvement opportunities, or to overcome inherent flaws of globalisation (e.g. Co-housing, Carpooling, Community-based agriculture, Micro-nurseries, Energy conservation and waste recycling procedures) – which we call "Social Innovation".

The analysis of consumers has to be updated according to some basic reflections about the psychology of the consumer, trying to avoid common misconceptions about an "average man" or a "model man" (Le Corbusier) with an homoge-

Chiara Cirinnà and Maurizio Lunghi (edited by), *Cultural Heritage on line: Empowering users: an active role for user communities. Florence, 15th - 16th December 2009* ISBN 978-88-6453-184-7 (print) ISBN 978-88-6453-187-8 (online)

neous rational mind; the audience has to be segmented and analyzed, bearing in mind that, along with a rational side, people's choices are driven by a powerful emotional thrust. Therefore, it's necessary to embrace the most advanced instances of the ethnographic research method (e.g. Filmed street interviews, psychoanalysis of the verbatim of online surveys, shadowing ...) to increase the understanding of dynamics and evolution of purchase processes and real-world location browsing. New technologies, moreover, allow the inclusion of a broader set of tools in the box of researcher: sensoristics and biometric analysis, fields that are centric for the development of non-canonical natural interfaces, is a valuable aid for the researchers willing to "read" between the lines of a consumer's choice.
Technology, though, may get in the way of consumers' needs satisfaction, if not correctly managed. Some philosophers (Paul Virilio, among others) strongly criticize the widespread technological positivism, stating that technologies have inherent costs implicit in their use, and incidents are integral to their very existence. Even if somehow catastrophic, this vision points out the strong influence that technological progress has on the boundaries of the traditional categories of thought, opposing, for example, reality to virtuality – indeed drawing complex and potentially dangerous scenarios. People are hit by these changes in several ways, the most prominent of which is what is generally referred to as *information overload.*
In fact, the ease of access to growing amounts of information – enabled by the Internet – mixed with the inherent flaws of the Internet as an ecosystem relying upon machines, brings a heap of unexpected side effects, ranging from the aforementioned copyright-bound issues, to the dangers tied to monopolies in information brokering.
The belief that the Internet can act as a collective memory has to be carefully reviewed, keeping in mind some issues: the information overload poses problems concerning information retrieval and the ability to control search tools; the risk derives from the fact that one big provider of search tools is capable to manipulate circulating information, and to decide what has to surface over the sea of the Internet;

- the ease of production of digital contents is contributing to the information overload by adding a lot of "noise", i.e. heaps of semiotic waste piling up throughout the Internet;
- the growth of low-cost consumer electronics industry introduced a risk factor connected with supports durability; while a clay tablet – if properly conserved – may last centuries, a low-end hard drive may fail in the arc of a week.

We propose, as an antidote to the situation described above, the widespread adoption of the methods traditionally belonging to the field of design – specifically Product Design. Companies like IDEO taught the world an important lesson: good design is to create experiences, not just products. This phrase is the key of a revolution that puts the consumer at the nexus of the design process, as only with the adoption of the viewpoint of the consumer, is possible to fully evaluate the performance of a given product or service. If the Key Performance Indicator is the customer satisfaction, the only measure of performance should be the customer himself.
The method that is emerging in today's design field involves three main features:

- Experience as the new paradigm in product and service design
- Storytelling as a tool to drive intra-psychic semantic processes
- Interface as a solution to reduce the complexity of the hyper-connected world

Experience is the key trait to understand today's dynamics of product value: the high cost, for example, of a coffee served over a table in Piazza S. Marco, Venice, is justified only considering the value of the whole experience, which comprehends the context in which it takes place; the only difference between products seemingly identical, moreover, is the experience they deliver – often through features not related directly to the product, for example the packaging.
A "good" experience, created by leveraging perceptive, cognitive, symbolic and semantic aspects, interacts positively with the customer's mind, with several interesting side effects. A natural attitude of humans – amplified by the continuous exposure to advertising – is to pursue the part of the world that promises pleasuring experiences – both sensually and intellectually: in a world overcrowded with information and stimuli, the attention is a limited resource; hence, the importance – for companies and designers – of the ability to catch it through positive experiences.
The focus on experience can help in achieving success because of its importance in mnemonic processes:

- a pleased customer will have the desire to repeat the experience; a loyalty bond with the producer (or provider) is forged;
- a pleased customer will share with friends and relatives his positive experience; the customer himself becomes a living advertising.

In short, a correctly designed experience allows designers to coherently leverage both the rational and the emotional side of the customer.
Core of the experience creation process, is the ability to tell a story in an engaging way, i.e. to acquire a storytelling proficiency. Novelists throughout history showed both that "only a great novel may be able to express the multiplicity of human experience, the paths of our interiority, the behaviours inside a society" (Edgar Morin), and that a novel draws its imaginative power from its ability to join distant concepts, images, things. The latter feature is what ultimately connects the activity of novelists and designers: the discovering of an unexpected proximity causes a pleasure in the discoverer that is rationally inexplicable. Physicist (and piano player) Victor Weisskopf wrote: «What is beautiful in science is the same that is beautiful in Beethoven. There is a fog of events and suddenly you see a connection.». As we can see, many sense generating activities (research, music, literature) share this relationship with proximity, distance and storytelling. We are "storytelling beings": narration allows us to better understand ourselves and others, and to build, accept and share experiences.
The user experience of a digital technology is undoubtedly tied to its interface. Basically a component that enables the use of the technology itself, the interface is a key element common to each and every tool, artifact or interactive system. The interface is only apparently the surface layer where information are exchanged and functions are triggered: it mirrors the deep structure of the underlying object and its ultimate purpose is to ease the interpretation and use through hints and suggestions (affordances, as Donald Norman calls them). This way the interface become an ubiquitous entity, ideally connecting the layer closer to the user with the designer's original intent, in a loopback process prolific enough to reveal new approaches to an existing technology. Said that, it's easy to understand why the Interaction Design field is so important as of today, and why the use of a given technology is so influenced by the nature of its interface. Social use of technologies, in fact, depends on how the interfaces expose and facilitate community building and interpersonal interaction. In general existing interfaces are the result of two opposed thrusts: one toward minimalism and feature subtraction, the other toward complexity and completeness; Jean Marie Floch, semiotic and advertiser, notes that these trends (recurring also in commercial and advertisement iconography) are directly connected to the two archetypical aesthetic patterns borrowed from architecture and art: the Classic and the Baroque. Traces can be found in the vast majority of graphic expression throughout recent history.
A field in which the production of "good" content and the ability to foster absorption and re-use of that content is rather important, is that of the so-called e-learning. Key to this process is the development – both as a concept and as a digital construct – of a digital-self capable to collect and rearrange the stimuli spotted by a given student/user.
The digital-self is, in our vision, a unique point of access to all online activities carried out by an individual, and to all the knowledge that person stored online. This structure, though, has to be constructed bearing in mind that – because of the risks described above – the more autonomous the digital self is, the more reliable it will be in the future: Internet, in fact, can provide neither a storage solid enough to guarantee continuous access to materials stored remotely, nor the means to efficiently organize and rearrange contents. An unmanageable digital self is useless, because it lacks the recombining features that make it a prolific structure: the recombination itself, in fact, help contents to stick to a user's memory.
In the e-learning process the relationship between recombination and recalling is very important. Every process of learning (reading, listening to a lecture, visit a museum, ...) should leave structurable traces in the "digital self", to be subsequently revised and reclassified: the reclassification of these "memory traces", decontextualized and disembodied, will follow the specific associative structures of the digital self, thereby increasing the "awareness" of semantic relationships and incremental accumulation of knowledge.
From the above reflections derive the innovative approach of the Experience Roma project. Experience Roma aims to bring innovation in cultural tourism by supporting the user/traveller all along his travel experience: from the planning, to the fruition, to the sharing of travel memories. The approach followed by the project is e-learning oriented and based on three main features:

- Segmentation of languages according to segmentation of target: each item residing in the database (POI, Routes, ...) is described using language in three different expressive styles (basic, historical and artistic, anecdotal) for the three types of users (general tourist, sophisticated tourists / specialist, junior user/ student);
- User generated content: the content generation system expects the participation, beyond site administrators, of the user community and a of group of brokers who select and enrich the contributions of users and prepare them for publication;
- Multimedia storytelling: contents are narrated through multimedia elements of various kinds, interactive maps, virtual laboratories and learning environments dedicated to deep studies.

The project, presented nationally – during a dedicated event in Rome, and through a book – and abroad – at the British Museum during an event related to cultural tourism – will be online in the course of 2010.

Increasingly cultural heritage institutions are providing access to their holdings and a variety of user services through their website. Feedback from users and potential users is essential to the sound development and ongoing improvement of these efforts. Professor Tibbo discussed the range of activities cultural heritage institutions can undertake to elicit such feedback in today's web environment including surveys, interactive websites, and web analytics tools. Her focus and examples explored such repositories serving higher education.

Chiara Cirinnà and Maurizio Lunghi (edited by), *Cultural Heritage on line: Empowering users: an active role for user communities. Florence, 15th - 16th December 2009* ISBN 978-88-6453-184-7 (print) ISBN 978-88-6453-187-8 (online)

In 1882, a young anthropologist from Washington, D.C., went west to collect objects for the Smithsonian Institution. He found a carved round shell, roughly eight centimeters across with two holes in it, buried in a small hill in St. Clair County, Illinois.

Confused about what the mysterious object was, the anthropologist brought it back to Washington and presented it colleagues in a salon for the exchange of ideas and knowledge. He took an isolated piece of history and placed it into a community that he hoped would illuminate it.

Over a century later I reproduced the anthropologist's activity digitally by presenting the same object online, to see what readers of my blog and my followers on Twitter could make of it, individually and by talking to each other. A different kind of analytical community: a virtual one.

I asked those following me online to work together to figure out what the object was. Participants in the experiment could post live comments on Twitter, and others could follow along by searching for the experiment's hashtag. (A hashtag is a hopefully unique string of characters that enables a search of Twitter to reveal all comments at a specific conference or on a particular subject.) I encouraged everyone to talk to each other and leverage each other's knowledge. And I gave the virtual community only one hour to solve the mystery of this carved shell.

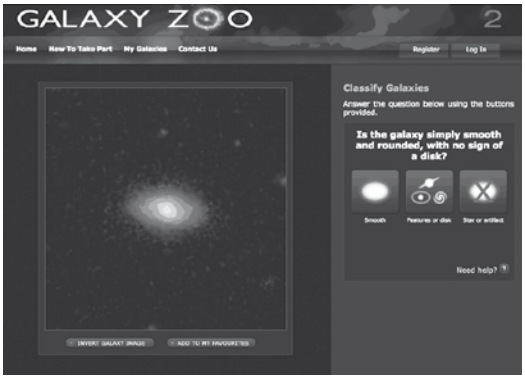
What happened along the way was as interesting as the result. First, Twitter was remarkably effective in spreading word of the experiment. Indeed, in the first five minutes about a dozen others on Twitter retweeted (rebroadcast) my historical challenge to their followers. This multiplier effect meant that within minutes many thousands of people heard about the experiment; over 1,900 actually viewed a facsimile of the shell on my blog in that hour.

Once the race was on, solvers took two distinct paths toward a solution. The first path was the one I was trying to encourage: some quick thoughts about facets of the object, followed by open, online scholarly debate. I mentioned that the object was made out of shell but was found far away from water in the Midwest (of the U.S.), which led to some interesting speculation about origins and movement of Native Americans, Europeans, and Africans. Others focused on the iconography of the spider; what could it symbolize and in which cultures was it used? These were decent lines of inquiry that one could imagine in the Victorian anthropologist's salon or in the back pages of an 1882 academic journal.

Incredibly, it took much less time than an hour for a solution: nine minutes, to be exact, for a preliminary answer and 29 minutes for a fairly rich description of the object to emerge from the collective responses of roughly a hundred participants. Solution: the object was an ornamental gorget from the Cahokia tribe.
Although my experiment was admittedly a bit of a stunt, I hope the implications of it are clear. When connected in the proper way to cultural heritage materials, virtual communities can be synthesized and catalyzed to produce helpful interactions and spawn new knowledge.

Those in the cultural heritage world should understand that virtual communities are already developing, often without our engagement. These communities, using services like Twitter, Facebook, and dozens of smaller, more focused outlets for social media, are self-organizing and often have a radical, do-it-themselves spirit. These virtual communities have no regard for traditional boundaries, such as institutional lines or barriers erected by academic credentialing. And they are currently working to create new functionality organically out of the way their particular communities function and exchange knowledge.
It is extremely useful for cultural heritage professionals to understand this new world enabled by social media and to think about how it fits in with the goals and mission of their work. Much can be gained by looking at some of the experiments that are going on online in this realm. I present here just a few examples, ones that I feel are most relevant to the work of museums, libraries, and universities.
Since I began with a discussion of my own experiment on Twitter, let me stay with that popular service for a moment. Jay Rosen is a professor of journalism at NYU who is very active on Twitter (http://twitter.com/jayrosen_nyu), providing what he calls a "flying seminar" to his many followers by pointing them to important publications and commenting upon current events. Although detractors see Twitter as nothing more than 140-character messages, Rosen has seen it as a place to build a community of discussion and debate. His activity on Twitter shows that social media is so flexible that it is what you make of it. Rosen calls what he does on Twitter "mindcasting" rather than "lifecasting" - that is, he uses Twitter to project and discuss what he is thinking about in his academic work rather than to inform others of his physical whereabouts or what is going on in his non-academic life.
Moreover, the use of Rosen's Twitter posts can be aggregated to get a good sense of what the journalism community is tracking at any one time. Rosen has set up a page called 40 Twits that shows which articles that he has highlighted are getting special attention by his readers.
Since Rosen set up his 40 Twits page, Twitter itself has launched features that make it easier to visualize the activity of a particular virtual community - to get the pulse of that community. For instance, I recently used Twitter's "list" functionality to bring together all scholars working in digital humanities (whether they are in universities, libraries, museums, or the technology field - again, across boundaries). This master list creates a unified stream of posts and makes it easier for scholars in this area to connect with each other.
In addition, the digital humanities Twitter list can, like Jay Rosen's Twitter followers, be used to provide a snapshop of aggregate interest. I recently set up *Digital Humanities Now* (http://digitalhumanitiesnow.org) to take the combined activity of the 350 people on my digital humanities Twitter list and show what they are reading and commenting upon. For those outside of the field, visiting the *Digital Humanities Now* site provides at a glance a good sense of what the discipline is interested in.
Cultural heritage practitioners should also look beyond the humanities for inspiration on how to use social media. A good example from the natural sciences is Galaxy Zoo (http://galaxyzoo.org), a website created by astronomers. These scientists currently are having to deal with a flood of information from telescopes and other sensors, and they obviously have limited personal time to sort through the millions of images and petabytes of data to find heavenly objects worthy of further study.
The astronomers behind Galaxy Zoo rather shrewdly decided to enlist the help of amateur astronomers and other enthusiasts of the night's sky to see what they could do together as a virtual community. Remarkably, more than 200,000 people have taken part in the community, helping to classify tens of thousands of galaxies. Doing aggregate work that no single astronomer or even a large group of professional astronomers could do, these "Zooites" have confirmed that galaxies do not prefer to spin clockwise or counterclockwise, and they have found unsual objects that astronomers have marked for more careful inspection. By curating their community well (and sharing

in the joy of discovery), the astronomers behind Galaxy Zoo have produced a valuable lession in crowdsourcing for academic purposes.

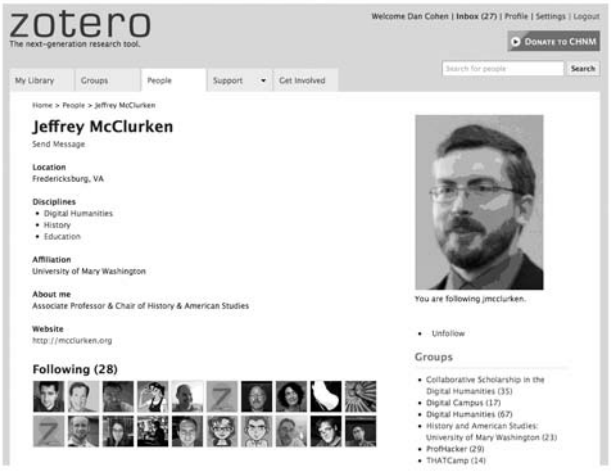


Museums, of course, have also engaged in their own crowdsourcing experiments. A project and associated software called Steve (http://www.steve.museum/) is being used by some museums to permit visitors to assign tags, or descriptive keywords, to works of art. The Indianapolis Museum of Art displays its Steve tags as another way to navigate its collection, and the community has now created scores of tags about virtually every aspect of IMA's holdings.

Unsurprisingly these tags sometimes differ from the controlled vocabulary and academic descriptions given to those very same artworks by IMA's curators. It is instructive for cultural heritage professionals to compare the outcome of Steve tagging and a museum's formal descriptions. For Robert Indiana's iconic "LOVE" painting, from the 1960s, the gallery label focuses on its geometry and frequent reproduction, the latter theme echoed by the amateur community in tags such as "omnipresent" and "postage stamp." But the community goes on attach some value judgments about the painting: "overexposed, overrated," as well as some helpful tags for those scanning the entire IMA collection for "word art."

What can we learn from these experiments in social media, crowdsourcing, and virtual community-building? First, there is still a yearning for validation and authority, which cultural heritage institutions are in a good place to capitalize on.

Second, social media is still an annoyance for many people, often because it appears as an unseemly mix of personal and professional life. (Thus Jay Rosen's urge to "mindcast" rather than "lifecast.") Finally, much of the work in this area is going on with the "secondary" products of scholarship rather than the primary ones. Virtual communities can be extraordinarily useful for the classification of galaxies or artworks, or adding context to a primary source; it is probably unreasonable to expect the crowd to produce a great novel together.
At the Center for History and New Media and George Mason University we have been engaged in our own experiment to build scholarly communities online through the exchange of bibliographical (i.e., secondary) information. The Zotero project, consisting of client and server software used by hundreds of thousands of scholars and researchers in virtually every field of endeavor in forty languages, has entered a 2.0 phase in which we are focusing on social media and the combined power of groups.
Zotero 2.0 allows researchers to establish online profiles and to expose information about themselves and what they are studying. As on Twitter they can follow and be followed by others. Should they choose to do so, they can share their bibliographies and associated scholarly materials.



A year ago we added the capability to create and join groups of any size and theme. These Zotero groups can be private or public, and to date nearly every discipline and subdiscipline is represented in thousands of macro- and microgroups. Some groups are loose networks of common interest; others are strongly curated by their members or their owners.
Regardless of how they are structured, these virtual communities are engaged in substantial exchanges of knowledge, and can be used for the creation of new collaborations and understanding.
Similar communities are just starting to arise within institutions and between institutions. At the City University of New York, the CUNY Academic Commons, which launched in 2009, attempts to unify online the many CUNY campuses through a common portal and the exchange of social media such as blog posts. HASTAC, the Humanities, Arts, Science, and Technology Advanced Collaboratory, brings together thousands of scholars working at hundreds of institutions to discuss the ways in which digital media and technology can transform their work and enable new forms of collaboration.
Social media can also be used by cultural heritage institutions as a live news feed to reach and engage a large audience. In the United States, the Library of Congress and the Smithsonian are both very active on Twitter, Facebook, and other social media outlets. The New York Public Library, perhaps ahead of its time, devotes an entire section of its website to its presence in social media and the ways that its audience can interact with it through those media.

Ultimately, these experiments in engaging and creating virtual communities lead one back to the objects of research and scholarship themselves. When I visited the Smithsonian as part of their "Smithsonian 2.0" effort to address social media and interactivity, I went behind the scenes and saw Abraham Lincoln's handball and paraphernalia from the presidential inauguration of Barack Obama. Just like the ceremonial Cahokia shell from the nineteenth century, these objects could have constellations of researchers and the general public gathering around them online. Cultural heritage institutions have always curated physical objects; in the twenty-first century they will also have to curate virtual communities around these objects.

What are computational methods in the humanities? What's the difference between using a computer and using computational methods?

Because the computer is a general-purpose modeling machine, it tends to blur distinctions among the different activities it enables and the different functions it performs. Are we word-processing or doing email? Are we doing research or shopping? Are we entertaining ourselves or working? But even though to an observer all our activities might look the same, the goals, rhetoric, consequences, benefits, of the various things we do with computers are not the same. I would bet that everyone here uses a web browser, a word-processor and email as basic tools in their professional life, and I expect that many of you are also in the humanities. Even so, you do not all do humanities computing – nor should you, for heaven's sake – any more than you should all be medievalists, or modernists, or linguists. However, if you are in any of these disciplines, one of the many things you can do with computers is to use computational methods, in which the computer is used as tool for modeling and analyzing humanities data and our understanding of it. Today, I simply want to point out that such activity is entirely distinct from using the computer when it models the typewriter, or the telephone, or the movie theater, or any of the many other things it can model.

There are any number of tools for modeling and analysis, depending on the nature of the source material: xml is a way of modeling text; mpeg is a way of modeling audio and video; GIS is a way of modeling geographic data, with other kinds of information layered on top of it; and we have various ways of modeling other kinds of information. The point, in each case, is that there should be some way of validating the model - some way of determining whether it is internally consistent, and some other way of determining whether it corresponds accurately to important features of the thing it models, even though the selection of those features and the importance given them will, inevitably, reflect the subjective interests and purposes of the person doing the modeling. Still, a model is a form of knowledge representation, and knowledge is always situated - in a person, and with a purpose - so, beyond accurately expressing those features of the object on which all observers can agree, the measure of success is not objective accuracy, but rather expressive completeness.

In addition to expressing the perspectives and purposes of the modeler, new perspectives on familiar materials can become available to others, as a result of the creation of digital primary resources. As an example here, I offer The William Blake Archive, which presents full-color images, newly transcribed texts, and editorial description and commentary, on all of Blake's illuminated books, with non-illuminated materials (manuscript materials, individual plates and paintings, commercial engravings, etc.) now coming on line. The Blake Archive makes it practical to teach Blake as a visual artist, by the simple fact of the economics of image reproduction on the web, and this is a fundamental change from the way I was taught Blake, through Erdman's text-only synthetic edition (which is also, by the way, available on the site).

There's a deeper impact of digitization, though, beyond increased access: that deeper impact is realized by those who do the digitization, provided that they are subject-area experts who are aware of the complexity of the source materials. In the act of representation, seemingly simple questions, like "is this poem a separate work, or is it part of a larger set of poems?" can be unavoidable - requiring some decision at the level of markup, for example - and they can also raise issues that are critical to understanding the work in question. However we may decide such questions, we are both informed and constrained by our own decisions, when subsequent and related issues arise. Likewise, with images, when we digitize, we choose file-type, compression, color-correction, and other settings based on what we consider valuable and significant in the image - and when our chosen strategy is applied across a large body of images, or when others come to our digital surrogate with purposes we hadn't shared or predicted, we are bound to confront the fact that our surrogate has been shaped by the perspective from which it was produced. In this sense, the real value of digitization for humanities scholarship is that it externalizes what we think we know about the materials we work with,

Chiara Cirinnà and Maurizio Lunghi (edited by), *Cultural Heritage on line: Empowering users: an active role for user communities. Florence, 15th - 16th December 2009* ISBN 978-88-6453-184-7 (print) ISBN 978-88-6453-187-8 (online)

and in so doing, it shows us where we have overlooked, or misunderstood, or misrepresented significant features of those materials.

No better example of this struggle between materials and intentions could be found, I think, than the documentation on the "Editorial Commentary " pages of the British Library's Nineteenth-Century Serials Editions project (http://www.ncse.ac.uk/commentary/index.html), which lay out the choice of materials, problems raised by multiple editions in serials, the construction of a "datamap" and a "concept map" for the materials, structural "segmentation policies," and the metadata schema that evolved during the course of the project team's effort to analyze and represent its six 19th-century serials. I'll quote just briefly from a now disappeared "work in progress" page that was once on the NCSE site, now no longer even in the internet archive, for its description of developing the NCSE datamap, in order to explain what I mean by this deeper impact of digitization. The datamap is a map of "data fields" in which the content of the NCSE primary materials will be represented, and it maps the relationships between those fields. Once an initial sketch of the map was prepared, it was tested against the primary sources in "a page turning exercise in which the team assimilated new data fields occurring in the source materials into the map and also reconfigured the map as appropriate." The team that went through this exercise noted that "this work required interpretation at every stage, our abstract conceptualisation of the source materials becoming increasingly concretely represented in the map as it was developed." Even so, the data don't always obey the map:

'The creation of the map has flagged up some potential challenges in the way in which our data might be rendered. As is evident from the map there are instances where relationships between its fields skip levels. (e.g. department items) and some items 'float' and can exist at almost any level (e.g. price). The dilemma facing ncse is thus whether to enforce an artificial framework upon the sources (top-down) or to attempt to adapt the framework to the sources (bottom-up).'

For me, this is very reminiscent of the exercise of developing the original SGML Document Type Definition for the Rossetti Archive, in doing which we went through an iterative process of modeling the components of Rossetti's paintings and poetry, an exercise that forced an explicit discussion of the nature of these materials, the relations between their parts, and the rules that could be deduced to govern the markup that would represent these. I guarantee that, in both of these cases, unless we had been digitizing the materials in question, and unless the scholar-expert had been party to that digitization, these discussions would never have taken place, and this explicit specification of the scholar's understanding of the materials would never have emerged. But these are the benefits of the early stages of digital humanities - the handmade phase, if you will, where the focus tends to be on scholarly editing as the analytic activity enabled by modeling the source material in digital form.

Beyond modeling, and beyond the hand-made phase of digitization, what does it mean to speak of computational methods? The word "method" implies a way of doing something; there should be something that can be computed on the basis of the representation, whether that's a matter of information retrieval, algorithmic transformation, statistical profiling or comparison - essentially, I would say "computational methods" involve some kind of analysis, and that analysis produces some kind of (reproducible) results. Those results are not, themselves, the end of the story: in the humanities, empirical results are most likely to be the beginning of the story - the evidence for an argument, the occasion for an essay, which still needs to be argued and essayed, in the same way we've always done.

What are the conditions that call for computational methods?

In the handmade phase, we could choose digitization, but we could choose not to digitize as well: scholarly editions, for example, can still be produced without digitizing the source materials. However, when we move from handcraft to industrial-scale digitization, we are required to consider computational methods in a different light. The primary condition that calls for computational methods is the availability of a large amount of data in digital form, with the possibility of reprocessing that data into other, purpose-built, representations. With respect to humanities research that focuses on text, we are certainly in that industrial phase: Google Books, as of October, had scanned about 10 million books. The HathiTrust, which is the shared digital repository that stores materials scanned out of the collections of some of the major research libraries in the U.S., had about 4.5 million volumes as of last month. Only some of this material is public domain, but the Google Books Settlement provides for the creation of at least two research centers that will provide access to the in-copyright material, for researchers in various disciplines who want to do "non-consumptive

research" with it (where non-consumptive" means, basically, that you're not supposed to be taking material out of the research environment).
As Franco Moretti points out, in Graphs, Maps, Trees, humanities scholarship normally focuses on a "minimal fraction of the literary field":
". . . a canon of two hundred novels, for instance, sounds very large for nineteenth-century Britain (and is much larger than the current one), but is still less than one per cent of the novels that were actually published: twenty thousand, thirty, more, no one really knows—and close reading won't help here, a novel a day every day of the year would take a century or so... And it's not even a matter of time, but of method: a field this large cannot be understood by stitching together separate bits of knowledge about individual cases, because it isn't a sum of individual cases: it's a collective system, that should be grasped as such, as a whole."
I think that what Moretti calls "the quantitative approach to literature" acquires a special importance when millions of books are equally at your fingertips, all eagerly responding to your Google Book Search: you can no longer as easily ignore the books you don't know, nor can you grasp the collective systems they make up without some new strategy—a strategy for using computational methods to grapple with profusion.
However, in order to exercise these strategies, in order to use computational methods, it is almost always necessary to be able to reprocess texts into new representations---transforming them, for example, into database representations, or indexes, or adding information about parts of speech, normalized spelling, etc. Particular purposes require particular representations, and different data-types will offer different features for analysis, but the basic point is that in order to do more than search and browse, it is almost always going to be necessary to reprocess the data, and one would usually wish to begin that reprocessing with the richest form of the source material.

What is the potential of such methods? What kinds of research questions can be addressed computationally?
When working with texts, computational methods can help us answer questions having to do with any number of empirical features of those texts and their authors, including vocabulary, syntax, grammar, sound, structure, reference, location, genre, gender, metaphor, intertextuality, and many other things. For example, we might examine

- historical trends in the use of language (for example, is there a golden age of the passive voice?)
- distinctive patterns of language that are characteristic of an author, by comparison to other authors of the same period (for example, what are the words that Jane Austen avoids, by comparison to her peers?)
- features that distinguish one genre from another, or one mode from another (for example, comedy vs. tragedy, or sentimentalism vs. realism)
- features that distinguish male from female authors in the same period
- the role of certain ur-texts, like the Bible, in shaping later texts
- authorship attribution, for example in multi-authored works like encyclopedias and so on.

Given the ability to reprocess texts, these questions can generally be answered at a level of specificity that would be impossible for a human reader to achieve, simply because the computer can keep track of empirical evidence at a very granular level. The role of the human interpreter is to understand and validate the methods by which the evidence is produced, and then to make sense of that evidence, in an argument.
Similarly, when working with other kinds of raw material—music, images, maps, 3D models, etc.—whatever empirical features that material offers will be available to computational methods, and those methods will support whatever meaningful questions can be asked on the basis of such features. Taking the whole process full-circle, one form of validation may eventually be production. In music composition, for example, computers have been able to learn algorithmic composition, using the features that characterize a particular composer well enough to produce new compositions that are plausible as works of that composer (see, for example Computers and Musical Style by David Cope, professor emeritus of music at the University of California at Santa Cruz; or listen to http://bit.ly/4QARWn for a couple of samples of the work of his program, named Emily Howell). In music, this is a matter of getting the syntax right; in language, this form of validation will be more difficult, because of the semantic component, but one day, we may see new works of fiction in the manner of famous now-dead authors, produced by computers. This is really just the generative inverse of analysis.

What has been the impact of such methods? Have computational methods changed the way we study and teach the humanities?

Certainly, both the hand-made and the industrial phases of digitization have had profound impacts on how we study and teach, or if they haven't, they should. The profusion of texts makes it all the more important that we teach students to understand the importance of editions, and to distinguish between reliable and unreliable editions. The presence of true scholarly editions in electronic form makes it possible to provide both students and researchers with unprecedented depth of access to the process and variety of artistic production (think back to the Blake example, or look at some of the scholarly editions produced by the University of Virginia Press—Melville's Typee with its manuscript, and an analysis of the process of revision that led to the final text. To take a different kind of example, there have been a number of interesting digital humanities projects based on correlating textual data with maps, for the purpose of analysis. I have first-hand experience with several of these, including the Valley of the Shadow project, which mapped military records and information from diaries and newspapers to produce interactive battle maps of some of the major campaigns in the American Civil War, and The Salem Witch Trials project, which mapped documentary records from the trial records to produce an interactive record of the location and spread of witchcraft accusations in the Bay Area colonies. In the case of the Civil War, these maps, combined with other data, helped to produce insights about the daily lives of individual during the civil war that no research to date has been able to match: read Ed Ayers's book, In the Presence of Mine Enemies to see exactly what kind of impact the coordination of very granular information from many different source can have, on the telling of history—this is a book that couldn't have been written without digitized primary source material. Ben Ray, in the Salem Witch Trials, was able to use the combination of maps and trial records to ascertain that a popularly held belief about the geographic concentration of accusers in one part of town, and accused in the other, was simply not true—and that, moreover, there were more accusations of witchcraft outside of Salem, in the larger colony, than inside it. These are some results that derive more or less directly from the act of digitization, by domain experts.

In my more recent experience, I've been working to develop tools that leverage digitized representation for the purpose of machine-aided analysis—in this case, text-mining. Over the last four years, I have worked with faculty, students, and computer experts at half a dozen different institutions in the United States and Canada, and at the National Center for Supercomputing Applications, to develop MONK, a workbench for text-mining across literary collections.

The full release of The MONK Project, available by authentication to about 50,000 faculty and 400,000 students at a dozen universities in the Midwest, includes about a thousand works of British literature from the 16th through the 19th century, provided by The Text Creation Partnership (EEBO and ECCO) and ProQuest (Chadwyck-Healey Nineteenth-Century Fiction), along with Martin Mueller's edition of Shakespeare (thirty-seven plays and five works of poetry), plus over five hundred works of American literature from the 18th and 19th centuries, provided by libraries at Indiana University, the University of North Carolina at Chapel Hill, and the University of Virginia.

MONK stands for Metadata Offer New Knowledge, and the metadata MONK provides is at the word level (part of speech, lemmata, position in the text, n-grams, etc.) for each of the 150 million words in this corpus. Behind the workbench interface, MONK's quantitative analytics (naive Bayesian analysis, support vector machines, Dunnings log likelihood, and raw frequency comparisons), are run through a toolkit developed at NCSA, called SEASR. Users typically start a project with one of the toolsets that has been predefined by the MONK team. Each toolset is made up of individual tools (e.g. a search tool, a browsing tool, a rating tool, and a visualization), and these tools are applied to worksets of texts selected by the user from the MONK datastore. Worksets and results can be saved for later use or modification, and results can be exported in some standard formats (e.g., CSV files).

In the process of designing MONK, we worked with humanities doctoral students and junior faculty who had specific research questions they wanted to answer, using these tools. For example, Sarah Steger was interested in sentimentalism in British fiction, and specifically, what distinguished sentimental from non-sentimental fiction, at the level of vocabulary—and, by extension, at the level of subject matter. She started by running naïve bayes routines on a training set of 409 mid-Victorian novels that she classified as either sentimental or unsentimental. The larger testbed, to which the software applied Sarah's training data, was 3,921 novels; ultimately, the software returned 1,348 chapters as sentimental, and it had detailed information about the language use that was characteristic of the sentimental chapters, and that distinguished them from non-sentimental chapters. She was able to get fairly definitive results on the words

that separate sentimental from unsentimental fiction, as well as learning that Dickens seems to be the archetype of sentimentality in this period of British fiction.

What are the limitations of such methods? What research questions cannot be addressed computationally?
In general, research questions that are wholly intuitive in nature, or that do not make use of empirical evidence in source material, will not lend themselves to computational methods. Aesthetic appreciations, likewise, don't benefit much from these methods. Arguments that depend on the performance of the critic may be assisted by evidence of the sort that these methods can provide, but they may not, as well.
Reflecting on our experience in the MONK project, where we based our analysis on meticulously prepared texts with in-depth linguistic information, my colleague and co-investigator Martin Mueller gave the following, fairly exhaustive, account of the limitations of our methods in the MONK project. He said,
"The computer has no understanding of what a word is, but it follows instructions to 'count as' a word any string of alphanumerical characters that is not interrupted by non-alphabetical characters, notably blank space, but also punctuation marks, and some other symbols. 'Tokenization' is the name for the fundamental procedure in which the text is reduced to an inventory of its 'tokens' or character strings that count as words. This is an extraordinarily reductive procedure. It is very important to have a grasp of just how reductive it is in order to understand what kinds of inquiry are disabled and enabled by it. A word token is the spelling or surface of form of a word. MONK performs a variety of operations that supply each token with additional 'metadata'. Take something like 'hee louyd hir depely'. This comes to exist in the MONK textbase as something like

hee_pns31_he louyd_vvd_love hir_pno31_she depely_av-j_deep

Because the textbase 'knows' that the surface 'louyd' is the past tense of the verb 'love' the individual token can be seen as an instance of several types: the spelling, the part of speech, and the lemma or dictionary entry form of a word."

Conclusion:
What's really changed? Well, perhaps nothing, for humanities scholarship that isn't primarily interested in modeling its source material in order to understand its structure or ontology, or scholarship that isn't especially interested in the evidence that is offered by that source material, for empirical arguments. But if your scholarship depends, to some extent at least, on empirical evidence, or if you are interested in the features that the computer can "understand," or if you are interested in correlating different kinds of evidence, along some shared dimension, then computational methods could change your work entirely, could lead to new answers to old questions, or even better, to altogether new questions. And as we approach, in the next decade, a time when all the books (not archives, but books) in research libraries are digitized, it may become harder to ignore the capabilities that computational methods offer to the scholar and the teacher. This is, in fact, how change has come to other disciplines—on the heels of a transformation of the bulk of their data from analog to digital. Our day is coming soon—in fact, it's already upon us, so it's time to begin thinking about how to cope with it.

I am delighted to be here and would like to thank the Fondazione Rinascimento Digitale for their invitation to speak at this special event. My talk will be different from preceding ones, since I will focus on convergence – specifically, Internet-driven convergence involving libraries, museums and archives. Is this an opportunity? Is it an inevitability? Or perhaps both? So today I would like to share some thoughts and examples with you about the changing nature of information and users in a digital age and how cultural institutions are responding, and how they will need to change in order to meet these challenges.

I believe that the collections of the future will integrate access to traditional print, museum and archival materials, with digital and three-dimensional objects, media formats onsite broadcasting and telecommunications. This extraordinary but achievable vision can only be realized if we bring libraries, archives and museums together through either physical convergence or through virtual convergence using the power of the Internet.

The role of technology

As we near the end of the first decade of the 21st century, we increasingly live in an age of connections, collaborations and convergence – between people, between professions, between countries, between ideologies. And technology helps foster the connections that make convergence possible, linking people and ideas across borders and boundaries. For libraries, archives and museums (LAMs), technology enables us to envisage collaborations and partnerships and projects that never would have been possible before the age of the computer, the Internet, digitization and so on. And technology allows libraries, archives and museums to bring together their collections and expertise in a way that can be accessed, ideally, from a single, central point by the user – who isn't concerned with our professional differences or demarcations, but simply wants quick and easy access to information.

This is quite a shift in mindset – and technology is a key factor behind such developments.

Most information searches begin with a search engine, and not with a library catalogue. Google – not JSTOR, or WorldCat, or any one of a number of other academic databases – is the default search option of choice for our users, regardless of what we, as information professionals, may think of this situation. And our users are no longer passive; they are using the information they find to create and share more information.

So we have all this content, all this technology, all these users. Yet we are almost drowning in the very information landscape we have created.

The challenge for libraries, archives and museums is to combine their expertise and offer solutions to such overload in a way that makes sense for all types of users all over the world.

However – it's also important to take a step back and realize that technology should not be viewed as the answer for everything. Technology should not drive digital projects. It is an enabler, but it is not an end-point.

Now, I'd like to look at the term "libraries, archives and museums" or LAMs. This is a kind of shorthand that lumps three closely related, yet distinct professions together – and while it may be convenient, we need to keep in mind that even as we use this term, the key commonalities and differences between the professions must be understood. If not, we run the risk of undermining trust and impeding serious dialogue across these communities.

So – how are we distinct? The key differences between libraries, archives and museums reside in the nature of their respective collections and in the culture and orientation of their professions.

Libraries have tended to focus on the description of metadata of books and serial publications, and the development of resource-sharing networks and practices such as interlibrary loans to serve users.

For archives, work has centred on establishing provenance and context, and the development of collective documentary approaches to establish the authenticity and meaning of archival documents and records.

Museums have focused their efforts on classifying or grouping individual objects based on their material, usage or

Chiara Cirinnà and Maurizio Lunghi (edited by), *Cultural Heritage on line: Empowering users: an active role for user communities. Florence, 15th - 16th December 2009* ISBN 978-88-6453-184-7 (print) ISBN 978-88-6453-187-8 (online)

some other common physical or social aspect, and the interpretation and use of objects in public exhibits designed to illustrate or explain a particular story, event, or other aspect of natural or human history.
So while the general public may not particularly care whether their information comes from a library, an archive or a museum, the professionals of these different institutions do. Again, a common understanding and appreciation of why each profession does things differently will determine the success or failure of convergence.
While LAMs are different in various respects, I think it's essential to remain focused on our commonalities. Libraries, archives and museums are all collection-based, they are service-oriented organizations, some more than others, and generally they are not-for-profit. They tend to have mandates that support learning, education, scholarly study and research, the advancement of knowledge, and the collection, preservation, organization, use and enjoyment of the collections in their care and custody.
These characteristics mean that LAMs are infrastructure-heavy. They need buildings that are big enough to house them, as well as micro-environments for various formats of material, and the security to accommodate collections, exhibitions, staff and public areas. They also share two other important features: a high demand for information technology and management, and skilled workforces to support collection management and outreach activities.
We all deal with some kind of evidence or object. The systematic discovery or protection of this evidence, the need to inventory and describe it, the problems of physical conservation, the importance of interpreting and presenting it for the general public, its value both in education and in scholarly study – these are issues we all face, and ones that bring us together. Comparing the solutions tried by different disciplines provides useful inspiration. Because all forms of evidence are equally important, contributing their unique insights to the mosaic that forms our social memory, our heritage.
So where are we at? What is the state of collaboration and convergence among the LAM community in today? To me, the answer is clear: the true convergence of libraries, archives and museums with each other is still in its infancy. In fact, if I may use an analogy, it is not unlike the silent movies of the early 20th century, when a moving image was projected on a screen, and a real, live orchestra provided the music and sound. Image and sound, two different media, were collaborating, but not yet fully integrated into an altogether new, organic whole, such as exist today.
On a regional, national or international level, a highly converged system or network of "libraries, archives and museums" today remains the exception, not the norm. Indeed, I think that in many cases, silos rather than synergies continue to define the LAM landscape. There may be a number of reasons for this situation. Appropriate levels of funding may not be available to foster effective collaboration and convergence. Individual institutions may be bound by mandates and statutes that set a restrictive set of operating guidelines and norms. The appropriate technologies may not be available or affordable. And contradictory cultures and attitudes at different institutions can pose enormous barriers. However, in spite of these obstacles, there are some exciting examples of convergence.
Library and Archives Canada was formed from the merger of the former National Library and the National Archives in 2004. Why? It wasn't for political or economic reasons. It was because separate institutions in the context no longer made sense. The two legal mandates were highly compatible, and the holdings were complementary. But most importantly the digital environment was blurring the distinctions between their holdings more and more, and the highly corresponding skills and competencies of the staff were not really being used to their full potential.
The lessons to be learned in any type of merger is not to drag the past into the future – that is, not to adamantly adhere to practices that may have worked well in the past, but no longer are suited to the realities of today and tomorrow. So – be flexible, be open, be willing to change, and take some risks.
Other examples of convergence of information resources:

- The Simthsonian Institution
- University of Calgary's Libraries and Cultural Resources
- Columbia and Cornell University Libraries Cooperation
- University of British Columbia Library provincial digitization project
- Europeana
- World Digital Library
- LAMMs at the International NGO level

I bring all these examples to light because I think what they demonstrate is that, yes, technology is a great enabler of convergence, of collaboration, of innovative partnerships that can take shape in many ways.
But what is more important is a willingness to recognize our commonalities. The more this kind of communication takes place, the more opportunities will present themselves, including new models of collaboration and convergence, new kinds of training and education, partnerships, preservation, collection management approaches, and search and online access offerings. Working together also lends LAMs a stronger and more articulate voice to advocate for funds and user rights, and to promote strengths and qualities.
However, there remain several challenges to advancing convergence or collaboration. There is obviously the technology challenge, always keeping up to date with new ways to create, share and preserve digital information resources. But I would argue that the biggest challenges lie more on the non-technical side, in the organizational and cultural attitudes that we have developed as information professionals.
Successful convergence across disciplines also demands a new set of diverse skills and competencies, people who combine subject-matter expertise with a knowledge of digital technology and information management. We need to create a new LAM professional, someone who can function in this increasingly cross-border landscape.
And if our institutions encourage such environments, this, in turn, can help us all attract and retain younger professionals – people who bring new skill sets, perspectives and attitudes to the table.

There is also a need to create guidelines for organizations that are working on converged collections, for example, ensuring that they are catalogued and accessible via a single website.
And finally, convergence and collaboration means that libraries, archives and museums will need to master the challenges of virtual exhibitions, multimedia, and fast and efficient digitization in order to take full advantage of the potential of digital technologies. We will need to become better communicators across many languages to promote joint projects.
The Web enables us to recall that the past was a holistic place – the documents, artifacts, paintings, books, historical sites, photographs – they all existed together, informed each other and collectively formed part of the holistic context of any historical action. The Web enables us to overcome the territorial boundaries that have arisen over decades – when we broke up the past, and put some in museums and some in archives, some in libraries, some in historic sites. The Web is all about convergence, and it enables us to overcome these boundaries and reassemble the past.
At the beginning of this talk, I posed the question: is Internet-driven convergence A) an opportunity? B) An inevitability? Or C) Both? The answer, as I see it, is probably both, although at this stage, more of an opportunity than an inevitability. Internet-driven convergence is a huge opportunity for libraries, archives and museums, and it is inevitable, given the relentless march of technology and the increasing interdependencies that are bringing us all closer together – whether we like it or not. It is our task, then, to take this technology and apply it in innovative ways, all the while remaining sensitive to our differences, confident of our commonalities and strengths, and focused on the users who will benefit enormously from our efforts if we are successful.

EU-funded research on digital libraries and digital preservation deals with leading-edge information and communication technologies for expanding access to and use of Europe's rich cultural and scientific resources. It also investigates how digital content created today will survive as the cultural and scientific knowledge of the future.

Three areas of research

The rapid pace of change of electronic devices and formats for recording, storage and use represents a threat to long-term accessibility of these resources. With the increasing proliferation of digital content this risk is imminent equally for businesses, the public sector and individual users. Digital preservation research aims at concepts, techniques and tools for ensuring availability of digital resources over time, while guaranteeing the integrity and authenticity of the information as originally recorded.

Leading edge technologies can enhance users' experiences with cultural and scientific digital resources. ICTs for capturing, rendering, modelling and visualising cultural artefacts support study and creative use of artefacts, and their aggregation into virtual collections. ICT-funded research establishes the basis for scalable and interoperable digital library platforms supporting digitisation and retrieval of heterogeneous content, in multimedia formats, from distributed collections and across languages.

These broadly are the three areas targeted by "Digital Libraries and Digital Preservation" research in the ICT programme:

- Digital preservation
- Digital libraries
- Cultural heritage

Target Outcomes

The ICT Work Programme 2009-2010, which is the basis for ICT Call 6, identifies six target outcomes related to the three domains, each associated to a specific funding scheme:

a) Scalable systems and services for preserving digital content, handling end-to-end workflows for different types of digital resources, guaranteeing their long term integrity and authenticity. The feasibility of solutions should be demonstrated in large scale testbeds (IP).
b) Advanced preservation scenarios:
   b1) Methods and tools for preserving complex objects, addressing the life-cycle of composite digital information instances (STREP).
   b2) Intelligent digital curation and preservation systems able to learn, reason and act autonomously, integrating tools and methods to support the complex decision making processes for appraisal, selection and management of diverse collections of digital resources. The system should ensure that the representation of the objects and their embedded semantic knowledge are preserved in order to support their future re-use. Appropriate verification scenarios should be integral component of the work (IP).
c) Innovative solutions for assembling multimedia digital libraries for collaborative use in specific contexts and communities, enhancing scholarly understanding and experiences of digital cultural heritage (IP).
d) Adaptive cultural experiences exploring the potential of ICT for creating personalised views of various forms of cultural expression, reflecting individual narrative tendencies (i.e. adapt to the background and cognitive context of the user) and offering meaningful guidance about the interpretation of cultural works (STREP).
e) Interdisciplinary research networks bridging technological domains (e.g. computing models, knowledge representation, visualisation and graphics), information and archival sciences, and social and cognitive sciences (NoE).

Chiara Cirinnà and Maurizio Lunghi (edited by), *Cultural Heritage on line: Empowering users: an active role for user communities. Florence, 15th - 16th December 2009* ISBN 978-88-6453-184-7 (print) ISBN 978-88-6453-187-8 (online)

f) Promoting the uptake of EC-funded research enabling the deployment of new ICT-based cultural and memory preservation services, leveraging the impact of associated national initiatives; identification of future 'Grand Challenges'; establishment of a pan-European network of living 'memory centres' for validations, demonstrations and showcases (CSA).

EXPECTED IMPACT

- Significant advances in the ability to offer customisable access services to scientific and cultural digital resources, improving their use, experiencing and understandings;
- Reinforced capacity of organisations to preserve digital content in a more effective and cost-efficient manner, safeguarding the authenticity and integrity of these records;
- Significant reduction in the loss of irreplaceable information and new opportunities for its re-use, contributing to efficient knowledge production;
- Strengthened leading edge research in Europe through restructuring of the digital libraries and digital preservation research landscape; leveraged impact of research results.

Research themes	Digital preservation	Digital libraries	Cultural heritage
Target outcomes in the ICT Work Programme 2009-2010	a) Scalable systems and services for preserving digital content (IP) b) Advanced preservation scenarios: - b1/ complex objects (STREP) - b2/ intelligent systems (IP)	c) Innovative solutions for assembling multimedia digital libraries (IP)	d) Adaptive cultural experiences (STREP)
	e) Interdisciplinary research networks (NoE) f) Promoting the uptake of EC-funded research (CSA)		

The table presents how the research themes are linked to the target outcomes and funding schemes.

AT A GLANCE:

Objective
ICT-2009.4.1: Digital libraries and digital preservation

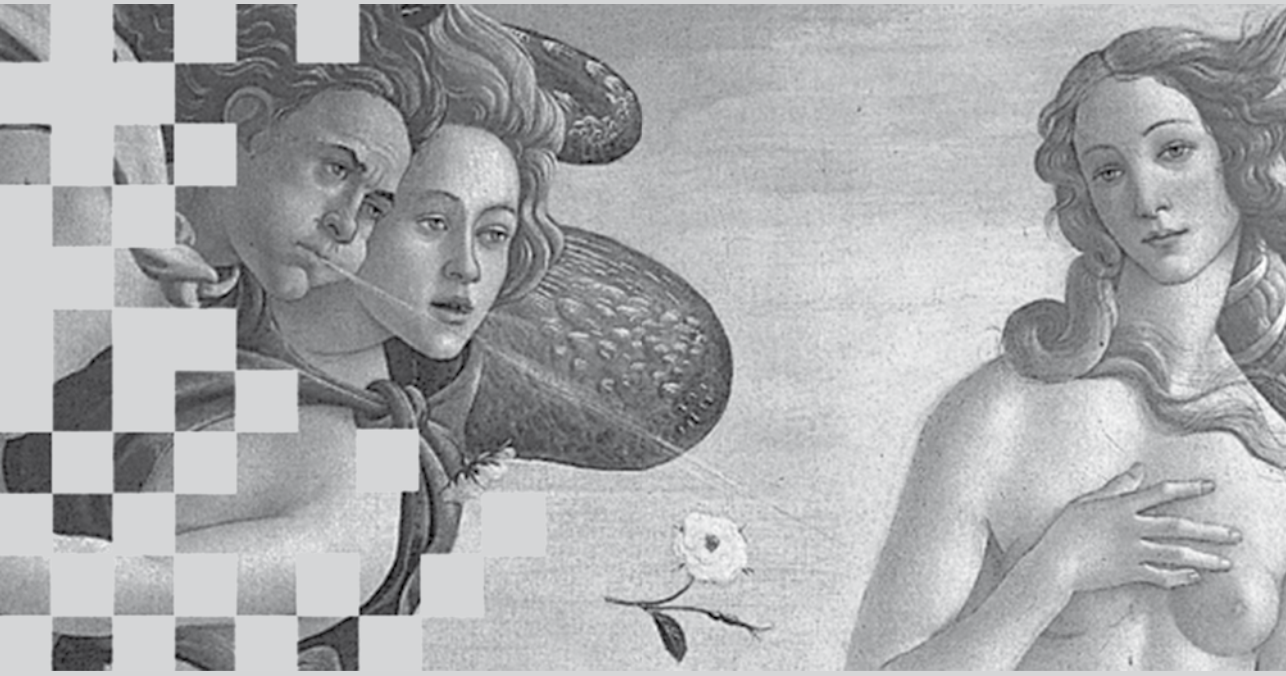
Funding schemes
- Small or medium-scale focused research actions (STREP)
- Large-scale integrating projects (IP)
- Networks of Excellence (NoE)
- Coordination and support actions (CSA)

Indicative budget
€ 69 million in total
- IP/STREP: € 56 million with a minimum of 50% to IPs and a minimum of 30% to STREPs
- NoE and CSA: € 13 million

Call
- ICT Call 6; submissions possible from
- 24 November 2009 to 13 April 2010

Web site
http://cordis.europa.eu/fp7/ict/telearn-digicult/digicult_en.html

Contact
infso-digicult@ec.europa.eu



Plenary Session
Invited lectures
Wednesday 16th December

Europeana is an online portal that contains 6 million digitised objects from Europe's museums, libraries, archives and audiovisual sources. It connects you to Europe's cultural heritage by providing access to items from over 1,000 heritage collections, including:

- Images - paintings, drawings, maps, photos and pictures of museum objects
- Texts - books, newspapers, letters, diaries and archival papers
- Sounds - music and spoken word from cylinders, tapes, discs and radio broadcasts
- Videos - films, newsreels and TV broadcasts

Some items in Europeana are world famous, from cultural institutions like the Rijksmuseum in Amsterdam, the British Library in London and the Louvre in Paris. Others are hidden treasures. The collection is continually being expanded and will reach 10 million objects in 2010.

From the beginning, Europeana has concentrated on creating a user-centric design. Focus groups, surveys, log-file analysis and user profiles are just some of the techniques used to ensure the site satisfies user needs.

Europeana is supported by funding from the European Commission and member states. Originally known as the European digital library network – EDLnet – it is a partnership of 100 representatives of heritage and knowledge organisations and IT experts from throughout Europe. They contribute to the Work Packages that are solving the technical and usability issues.

The project is run by a core team based in the national library of the Netherlands, the Koninklijke Bibliotheek[1]. It builds on the project management and technical expertise developed by The European Library[2], which is a service of the Conference of European National Librarians[3].

Overseeing the project is the EDL Foundation[4], which includes key European cultural heritage associations.

[1] http://www.kb.nl
[2] http://www.theeuropeanlibrary.org
[3] http://www.nlib.ee/cenl/about.php
[4] http://version1.europeana.eu/web/guest/edl-foundation

Chiara Cirinnà and Maurizio Lunghi (edited by), *Cultural Heritage on line: Empowering users: an active role for user communities. Florence, 15th - 16th December 2009* ISBN 978-88-6453-184-7 (print) ISBN 978-88-6453-187-8 (online)

My paper intends to explain the strategy at national and European level adopted by the Italian Ministry of Cultural Heritage and Activities in order to coordinate digitisation initiatives and to promote access to cultural information.
The Italian Ministry's strategy is focusing on the integration of existing information systems, on the recovery of databases not in line with the current international standards, the creation of web sites and of cultural portals.
The Central Institute for the Union Catalogue of Italian Libraries and for Bibliographic Information (ICCU), directed by me, is the institute coordinating all cataloguing and documentation activities carried out by Italian libraries.
The ICCU promotes and develops programmes, studies and scientific initiatives concerning cataloguing, inventories and digitisation of the bibliographic and documentary heritage preserved in State libraries and other Italian public and private institutions.
The Institute coordinates the development and dissemination of the librarian cultural heritage in order to define a national system of services.
Through the Technologic Observatory of Cultural Assets and Activities (OTEBAC) the ICCU also promotes the harmonisation of digitization standards and the management of digital resources across all sectors of cultural heritage.
The ICCU:

- coordinates, promotes and manages the catalogue, the newtwork of the National Library Service (SBN), the Inter-Library Loan services and documents' delivery;
- coordinates, promotes and manages the national databases concerning the census and bibliography of manuscripts, antique books and the Italian library database;
- manages the procedures of local applications in accordance with the National Library Service:
- promotes and coordinates the production of national regulations and the dissemination of international standards and cataloguing rules, assuring the uniformity of the catalogue and the production of the bibliographic control;
- participates at international level in the production and updating of bibliographic standards and formats;
- promotes and coordinates the production of standards and guidelines for digitisation of bibliographic and documentary heritage, with respect to the phases of archiving, management, conservation and access to digital resources;
- coordinates the monitoring of digitisation projects and oversees the publication and dissemination of digital resources, integrating them with SBN;
- manages education and training activities in its own areas, offering traditional and e-learning courses;
- takes part into international projects concerning the dissemination of information and digitisation of cultural and scientific heritage, such as ATHENA, CERL, DC-NET, DPE, MICHAEL, TEL, Europeana;
- carries out editorial activities.

The activities, researches and technical actions are promoted by ICCU in accordance with the general directives of the Ministry of Cultural Heritage and Activities and of the General Directorate for Library Heritage, Cultural Institutes and Copyright.

What follows is a summary presentation of the most important projects monitored by ICCU.
Internet Culturale: digital catalogues and resources of Italian libraries
(http://www.internetculturale.it)
Internet Culturale is a multilingual portal (in Italian, English, French and Spanish) that allows users to access documents and digital resources of Italian libraries and to find information on their activities.
Through the portal users can:

- search within the national online loan service;
- visualize and download images of linked digital collections;

Chiara Cirinnà and Maurizio Lunghi (edited by), *Cultural Heritage on line: Empowering users: an active role for user communities. Florence, 15th - 16th December 2009* ISBN 978-88-6453-184-7 (print) ISBN 978-88-6453-187-8 (online)

- access to the InterLibrary Loan service;
- search in the digitised historical catalogues of Italian public institutions;
- search in the Italian libraries database.

In the section devoted to special catalogues users can find information on:
- Italian editions of the 16th century (EDIT16);
- manuscripts in Latin alphabet owned by the libraries participating the Manus national census;
- the bibliography of manuscripts produced by the Institutions joining the BibMan project;
- the description and digital reproduction of Greek palimpsests.

Internet Culturale also allows browsing, with the help of three-dimensional technologies, of animations, hypertexts, through virtual exhibitions dedicated to famous characters (e.g. Svevo, Verdi, Totò) and tourist cultural itineraries or gourmet journeys.
Internet Culturale also devotes a large section to musical heritage allowing the search by institutions, projects, collections, authors and by rare items such as autographs (there are for example 272 items on Donizzetti, 70 items on Puccini, 25 items on Bellini), by musical itineraries.
Contributions come from State Libraries and Monumental Libraries, Libraries of Conservatories, Museums, Civic Institutions, State Archives, and important musical institutions.
The digital objects are almost all preserved in the ICCU repository with a number of approximately one and a half million images that form one of the richest online database.
The digitisation of the historical catalogues of libraries took place in the framework of the Italian Digital Library Project and includes the scanning, in image format, of over 200 Italian public library catalogues, making available about 7 millions images accessible through Internet Culturale.
This collection gathers digital reproductions of catalogues, in cards or in volumes, organised in different ways (alphabetic, topographic, systematic, mixed) and relative to various materials (printed editions, manuscripts, special materials). The headings were taken in the form in which they appear in the catalogue (without standardization) in order to respect the "historical authenticity" of the catalogues.
Information about libraries and about each digitised catalogues (history, organization, special ways of searching) are also available.
Internet Culturale is integrated with CulturaItalia, the aggregator of contents and access point to databases, websites and digital collections involving all sectors of cultural heritage.

Cultura Italia (http://www.culturaitalia.it)
CulturaItalia proposes a guided access to the world of Italian culture. Thanks to innovative IT solutions, the portal gathers and organises millions of information-related elements on the resources that make up the country's vast cultural universe, and makes them available to all web users.
The data on cultural resources are not produced by the Portal, but are provided directly by the entities who own and manage the resources. All the operators in the cultural sector – public administration and private companies – upload only the "metadata", that is, the descriptive information of the resources in their possession. CulturaItalia offers to users the opportunity to consult and search information on Italian cultural resources within a single site.
The user, through the Portal, accesses a database of "metadata" which gathers and organises the information arriving from all providers participating in CulturaItalia. Users can discover resources of all types which make up the country's extensive cultural heritage (museums, photographs, libraries, archives, galleries, exhibits, monuments, videos, discs, etc.), can carry out searches for scientific research or just out of simple curiosity. CulturaItalia is an "open" system in that it grows and develops in sync with the new information contained in the resources that enrich its database. The Portal does not, per se, contain resources on the Italian cultural heritage, but rather proposes itself as a starting point for a guided search towards other sites.
The Portal offers a service to users, who will have at their disposal a single location from where to start their own search itineraries online, in terms of Italian culture, and the operators in the field, who can take advantage of a high-quality

showcase to promote their own contents. Once the resources of interest are located in CulturaItalia, the user can consult them directly at the data source, by heading to the provider's site or by contacting them via other channels, to complete their process of analysis and understanding.
The Portal is an answer to the needs of an expert public as well as to the needs of the general public. CulturaItalia offers to specialised users, such as students, researchers, and those employed in the cultural sector, the opportunity to carry out targeted searches that correspond to very specific interests using a very advanced software. For non-specialised users, such as citizens and tourists, the Portal can stirr curiosity and offer opportunities to discover or find out more about cultural resources available in the territory, mostly thanks to its editorial content (thematic itineraries, articles, highlights, events, columns) published to spotlight the stores of "metadata" present in the website database.
The project is promoted and managed by the Ministry of Cultural Heritage and Activities (MiBAC) with the scientific consultation of the Scuola Normale Superiore of Pisa. CulturaItalia is supported by many organisations and institutions that belong to the world of Italian culture and provide the relevant information, that is to say the real "primary resource" of the Portal. Precisely because of its ability to integrate in one system the information-related elements of many different entities, CulturaItalia is a leading project in Europe and has been used as a reference for many other countries hoping to promote similar initiatives.
CulturaItalia is the national aggregator which feeds Europeana, the European Digital Library.
Summarizing, CulturaItalia aggregates contents at Italian level, while Europeana aggregates contents at European level. Both projects use the same standards and share the same approach and philosophy. Both projects involve all heritage sectors.

MICHAEL, Multilingual Inventory of Cultural Heritage in Europe, <http://www.michael-culture.org>
The MICHAEL and MICHAEL Plus projects were funded through the European Commission's eTen programme, to establish a new service for the European cultural heritage.
The MICHAEL project was a partnership between France, Italy and the UK to deploy a cultural portal platform that was developed in France. MICHAEL Plus then extended the MICHAEL project to the Czech Republic, Finland, Germany, Greece, Hungary, Malta, the Netherlands, Poland, Portugal, Spain and Sweden. The two projects were closely aligned.
The projects focussed on the integration of national initiatives in digitisation of the cultural heritage and interoperability between national cultural portals to promote access to digital contents from museums, libraries and archives.
The projects have established this international online service, to allow users to search, browse and examine descriptions of resources held in institutions across Europe. We hope that the technical standards and sustainability model that we have established for the project will mean that more countries will provide their contents to the portal in future times.
Through the multilingual MICHAEL service people are able to find and explore European digital cultural heritage materials by using the Internet.
MICHAEL's objective were:

- A European cultural heritage inventory, available to all and providing access to cultural heritage resources.
- Sustainable management for the project to continue.
- Endorsement and implementation at a national government level, in order to underpin further funding as required.
- A methodological and technical platform, which makes it easy to add new national instances of MICHAEL, thus improving the content and user bases.

The technical results of the MICHAEL project can be listed as follows:

- The MICHAEL data model for multilingual digital cultural heritage inventories
- An open source technical platform for national instances built on Apache Tomcat, Cocoon, XtoGen, XML etc.
- Interoperability protocols for national instances to contribute data to the European service
- European MICHAEL search portal
- Methodology and model which is easy to deploy in additional countries.

MICHAEL currently includes some 10.000 digital collections from 4000 cultural institutions in Europe representing millions of di data.

The MICHAEL Consortium (19 states and 40 partner) has set up an international association known as MICHAEL-Culture AISBL, to grant sustainability and to allow for further service developments. MICHAEL-Culture is a member of the European Digital Library Foundation, that manages and monitors the implementation of Europeana.
The Italian instace of MICHAEL is interoperable with CulturaItalia.

When considering experiences at European level, I wish to illustrate two other projects coordinated by ICCU.

ATHENA (Access to cultural heritage networks across Europe)
(http://www.athenaeurope.org)
In November 2008, the 'Network of Best Practice' was launched. This is a new project known as the eContentplus programme developed by the MINERVA network.
Its partners come from 20 EU Member States together with 3 non-European observers. 109 major museums and other cultural institutions are directly associated with the project, and 20 European languages are represented. It is coordinated by the Italian Ministry of Cultural Heritage.
ATHENA's objectives are to:

- Support and encourage the participation of museums and other institutions not yet fully involved in Europeana;
- Produce a set of scalable tools, recommendations and guidelines, focusing on multilingualism and semantics, metadata and thesauri, data structures and IPR issues. These will be used by museums to support internal digitization projects and to facilitate the integration of their digital content into Europeana;
- Identify digital contents that are present in European museums;
- Contribute to the integration of the different sectors of cultural heritage with the overall objective to merge all these different contributions into Europeana. This will be carried out in cooperation with other projects more directly focused on libraries and archives;
- Develop a technical infrastructure that will enable semantic interoperability with Europeana.

ATHENA will:

- Bring together relevant stakeholders and content owners from all over Europe;
- Evaluate and integrate standards and tools for facilitating the inclusion of new digital content into Europeana;
- Enable the user of Europeana to have a complete experience of European cultural heritage;
- Work with existing projects (Europeana, and Michael are both present in ATHENA);
- Develop links and joint activities with other relevant projects in the Europeana 'cluster' (for example EuropeanaLocal).

DC-NET - Coordination action contributing to European Research Area Network (http://www.dc-net.org)
DC-NET is an ERA-NET (European Research Area Network) project, financed by the European Commission under the e-Infrastructure - Capacities Programme of the FP7.
The main aim is to develop and to strengthen the co-ordination of the public research programmes among the European countries, in the sector of the digital cultural heritage.
This scope will be pursued by the participating Ministries of Culture by endorsing a Joint Plan of Activities, to be initiated already during the project time-frame, through a wide and intensive programme of seminars, workshops, meetings and Presidential conferences dedicated to the encounter of the digital cultural heritage sector with the technological research and the e-Infrastructure providers in Europe.
The main objective of the DC-NET project is to develop and to strengthen the co-ordination among the European countries of public research programmes in the sector of the digital cultural heritage. The project will integrate the research capacities of the participant member states, will identify their communalities and will valorise existing programmes and projects in order to initiate the deployment of a wide and comprehensive European e-Infrastructure that will increase the research capacities of the digital cultural heritage community.
The DC-NET project will contribute to the coordination of the research priorities of Ministries of Culture, their Agencies and other cultural bodies (museums, libraries, archives, audiovisual, archaeological sites, etc.) across Europe in the

area of the eInfrastructures targeted to the digital cultural heritage. The project will coordinate the manner in which cultural actors can and should engage with national and European e-Infrastructures to generate innovative services, tools and data sets to support the research of multidisciplinary communities. A programme of seminars, workshops, meetings and conferences will involve all the relevant stakeholders. A plan of joint activities for e-Infrastructure-enabled research in the sector of digital cultural heritage will be generated and the joint activities will be initiated.

ICCU's backoffice in the field of digitisation currently is mainly devoted to three activities:

- Development of the campaign "Join CulturaItalia" targeted to cultural institutions.
- The workflow consists in five main phases: a) Identification of databases to be harvested; definition of the agreements between MiBAC and the cultural institution; 3) definition of the data amount and level of description to be made available to CulturaItalia; 4) the metadata harvesting; 5) the quality check.
- Development of institutional cultural websites

MiBAC is continuing to diffuse all guidelines and publications produced by the MINERVA project in the field of cultural websites.
Furthermore, it promotes "Museo & Web", the open source content management system developed by MiBAC in the framework of MINERVA.
Currently MiBAC institutions have developed more than 300 websites; half of them made with Museo & Web.
In the last years, institutions changed the way they perceive websites.
According to Maria Vittoria Marini Clarelli, director of the National Gallery of Modern Art (GNAM), a museum's website represents a tool to prepare/complete the visit; a way to facilitate the cultural mediation; a free of charge way to access; a window. Feedbaks proved that users expect that an institutional website should be trustworthy and reliable, that contents should be maintained up to date, that the website should grow in time.
A well done website changes the institution's way to communicate; produces more service and qualitative information; stimulates more internal communication and coordination, as well as benchmarking with other institutions.

- Supporting institutions in the activities of digitisation

As regards digitisation, there are still several open issues: high costs, insufficient know-how (with the exception of librarians), insufficient knowledge of resource description standards and harvesting protocols, a general need of training, often a low quality of metadata, products available only off-line (CDs-DVDs), a generalised used of proprietary software.
Currently ICCU set up an Italian cross-domain working group for the harmonization of ALM sectors metadata standards in a cross-domain perspective. The Working group is composed of ICCU experts working together with the Central Institute for cataloguing and documentation (ICCD) and the General Directorate for Archives of the Italian Ministry of Cultural Heritage.

But, another real barrier to the free dissemination of contents are Copyright and Intellectual Property Right (IPR). This issue is common to several European Member States.
MiBAC, in accordance with the European strategy, will soon adopt 'Creative Commons' licences in the framework of CulturaItalia.
The European project ARROW (http://www.arrow-net.eu), coordinated by AIE (Associazione italiana editori, Italian Publishers Association), is trying to clarify, at a European level, issues concerning orphan works and out of print works, in the field of publishing.
As far as multilingualism is concerned, the barrier of European linguistic diversity is dramatically evident regarding common access points to distributed databases. Investments are insufficient and tools available are not yet suitable.
Partnership with privates is another issue to be investigated. It could be a way to guarantee the online presence of cultural works under copyright (books) or intellectual property right (i.e. contemporary artworks). Of course agreements must be checked carefully by the institutions involved. In fact, after Google's recent digitisation proposals to several

European libraries, national governments decided to share a common position in order to guarantee public access and reuse of contents susceptible of Google digitisation.
On the contrary, many efforts are carried on in the field of digital preservation. ICCU made digital repositories available for institutions, while the Central National Library of Florence, the Central National Library of Rome, and the National Library of Venice are working in close cooperation with Fondazione Rinascimento Digitale at the initiative Magazzini Digitali ((Digital Repositories) <http://www.rinascimento-digitale.it/magazzinidigitali.phtml>.
An important aspect to be taken into consideration are users. Institutions are used to consider the users of cultural websites, digital libraries and portals as ALM professionals (librarian, archivist, curators, etc.) scholars and experts, information
scientists, students, tourists, creative or just curious people.
Institutions should take the users' age and the distinction between digital immigrants and digital natives into account.
Digital immigrants are people who are "old" enough to have attended the world of "yesterday" but "young" enough to have lived in the world of "today". So far they played a central role in knowledge and culture building and in developing the Internet. Generally speaking, up to now, websites, digital libraries and cultural portals have been conceived by digital immigrants.
Digital natives are born after 1980. They study, work, write and interact in a different way from their "antecedents"; they read blogs instead of newspapers; often they only get to know each other on the Internet; they have never written a letter, but they chat or text friends; many of them have never got in a library, they download music illegally, believing that it's legal; they handle easily and skillfully digital contents creating new ones. Summarising, the main aspects of their lives are mediated by digital technologies.
Is the institutional web ready to satisfy the digital natives'needs? It is necessary to monitor and analise the audience behaviour in order to get responses to this issue.
Recently CulturaItalia launched an online survey to measure the users' satisfaction, based on the Minerva Handbook on cultural web user interaction. The results will be published in spring 2010.

I wish to end my speech bringing at your attention a successful story: the experience of SBN, the Italian library network.
SBN (http://www.sbn.it - Servizio bibliotecario nazionale) is an infrastructure of national services for users promoted by the Ministry of Cultural Heritage and Activities, coordinated by ICCU and acting in cooperation with Regions and Universities.
State libraries as well as local, university and private libraries operating in different sectors are linked to the SBN.
The libraries participating in SBN (approximately 4000) are organised in 69 Nodes distributed across the national territory, connected to a central system (the Index), which sets up the general catalogue of libraries belonging to the network.
In 2002, with the launching of the Index evolution project, the rationalisation, integration and renovation of the Index central database was developed and the opening to other systems and management of different levels of cooperation was also envisaged.
The main aims of the project were:

- the technological renewal of hardware and software
- the opening of the SBN Index to management systems of a non-SBN library using the most widespread bibliographic formats (UNIMARC, MARC21);
- the management of diversified levels of cooperation;
- the development of new activities such as for example derived cataloguing and cataloguing of special materials;
- the development of government and monitoring functions of the system and of the increase of the databases.

Since 1997 the Index has been available to users through the OPAC (Online Public Access Catalogue) system that allows access to the contents of the catalogue, with research methods that are user-friendly and articulated, and to use its connected services.

The OPAC database currently contains about 11 millions of different pieces of information and 50 millions of localisations; it can be consulted through two interfaces (http://opac.sbn.it and http://www.internetculturale.it) with approximately 42 millions searches carried out in 2009.

Other important features of the OPAC SBN system across the Web are:

- the service of SBN ILL Interlibrary Loan;
- the integration with the local OPACS;
- the access to the cards of the identified libraries ;
- the presentation of search results in various formats among which UNIMARC and USMARC formats;
- the UNIMARC export of individual bibliographic items;
- the possibility of search and presentation for authority entries regarding authors included in the SBN Index (currently 36.000);
- the possibility to operate as a Z39.50 client and therefore to interrogate other Z39.50 catalogues at national and international level.

All the Italian initiatives mentioned in my speech have a common distributed approach, share coordination structures at local, regional and national level, have close links with national digitization strategies, benefit from the active participation of hundreds of cultural institutions (at all levels and in all sectors) and, last but not least, have in general a cross-domain approach with museums, archives and libraries.
All these efforts have generated a number of benefits for all involved stakeholders: local administrations, institutions and end-users.
The modalities to access information have exponentially increased, interoperability is undergoing rapid changes, local initiatives are enhanced thanks to new scenarios.
But many efforts are still necessary: more coordination and cooperation among stakeholders; more training; advocacy campaigns for funding; more cooperation among cultural institutions and research institutions.
At European level, the new Commission ambition is to realise a European digital agenda including a legislative action. The aim is to create in Europe “a modern and legal platform, competitive and “consumer friendly” for a unique market of online creative content”.

During the second half of the 20th century, in Italy as well as in many other countries of the developed world, the archival landscape? underwent deep changes that helped reshape its territorial organization and the balance between its various institutional components. At the base of those changes are some typical phenomena of contemporary society on which archive literature has widely dwelled: the advent of new historiographic trends, addressed at documentary sources that were generally neglected or undervalued up to that time; the interest in recovering individual and collective memories, as well as local and territorial traditions; consequently, the new attention in archives as vehicles of memories and traditions and the rise of new expectations in them from an ever-growing, culturally and socially heterogeneous public; the increasing importance and activism of local and regional institutions and of various cultural and social organizations who headed up a crisis in the traditional balance between central and local powers, favoring the second as opposed to the first.

In Italy the outcome of these changes was the gradual exhaustion of the historical records concentration and keeping model established after the Unity, which was based on the centrality of a network of provincial State Archives, directed by State central administration located in Rome. A different model, gradually affirmed for the past few decades, is based on the coexistence of many archival holders (archives of municipalities, provinces, regions, cultural institutions, research centers, business archives, etc.) and on the development of various initiatives for collecting, describing and promoting archives, supported by a large number of public and private bodies.

This dissemination of archival institutions and initiatives presents many problems of connection and coordination and urges archivists and State, regional and local administrators to establish new forms of collaboration, which are particularly important in a phase like the current one where financial resources continue to decrease and the risk of dispersing them without lasting outcomes is increasing. Therefore, as claimed by many parts, it is urgent developing sustainable models of archival polycentrism capable of improve its positive features, while limiting the most critical, not to say negative, ones, such as an excessive dispersion and fragmentation of archival institutions and projects, which probably users of archives find unreasonable and which, at the same time, risk to put in danger the preservation of the extremely rich and precious Italian archival heritage.

The Second National Conference of Archives, held in Bologna from 19 to 21 November 2009, was dedicated to the need to "make a system" for governing the archival polycentrism. It was an important moment for debates and proposals where representatives from State archives administration, regions, local authorities, cultural institutions and other public and private organizations discussed the need for coordination and cooperation on different fronts: from the construction of common repositories for archives to common strategies for digital preservation, from archival education to initiatives for promoting and publicizing archives. The result was the design of a National Archives System (SAN) to be built in the near future with one's own national government bodies, regional coordination committees and common repositories and archival services at the local level[1].

The National Archives Portal and the Archives Resource Catalog

The core component of the National Archives System will be a web portal, which has momentarily been given the name Portale Archivistico Nazionale (National Archives Portal) or PAN, which should present itself as the integrated access point to national archival resources, irrespective of the juridical status and affiliation of the institutions or organizations that developed them. The Portal, whose planning and realization began in 2009 by the Directorate General for Archives

[1] See the Conference web site <http://www.conferenzanazionalearchivi.beniculturali.it/index.php?it/1/home>, where all the preparatory documents and the final one can be accessed.

Chiara Cirinnà and Maurizio Lunghi (edited by), *Cultural Heritage on line: Empowering users: an active role for user communities. Florence, 15th - 16th December 2009* ISBN 978-88-6453-184-7 (print) ISBN 978-88-6453-187-8 (online)

of the Ministry for cultural heritage and activities, will be divided into multiple sections and will carry a lot of complex contents, such as:

- an "Archipedia", which is a sort of encyclopedia of archive definitions and concepts compiled in cooperation with the portal's users and including a glossary which will provide simple and concise explanations of archival technical terms for novices and unskilled users ;
- a database of bibliographic resources related to Italian archives;
- research guides, virtual tutorials and other materials for different typology of audience (for example: teachers, students, genealogists, historians, etc.), including novices and unskilled ones;
- specific thematic sections or sub-portals dedicated to describe and make available different typologies of records (business, genealogical, cartographical and fashion archives etc.);
- editorial and multimedia contents, including news, virtual exhibits, photo galleries and so on and so forth for illustrating the multiple aspects of the world of archives and records;
- a digital archive that allows access to digital reproductions of fonds and series published on web sites of local, regional or national archival institutions;.
- Web 2.0 tools for communicating with the users of the portal, allowing them to collaborate in creating its contents and offering them the possibility to build communities on specific topics and research projects.

The central component of the PAN will be the Catalogo delle Risorse archivistiche (Archival Resource Catalog) or CAT that proposes to be a coordinating and integrated access point to archival descriptions stored on archival databases and systems developed at the regional and local level, respecting their autonomy and specificity.
In fact, one way that the archival polycentrism mentioned above has been manifested in Italy, was the development of many software and systems for producing digital descriptions of archives and for publishing those descriptions on the Web. Systems have been developed at the national level by the State archives administration (such as the General Guide to the State Archives[2], the State Archives Information System or SIAS[3], the Unified System of Archive Supervising Agencies or SIUSA[4], the Mediterranean Historical Multimedia Archives[5]; by some provincial State Archives (the State Archives of Florence[6], Milan[7], Bologna[8], Naples [9], Venice[10]; by some regions (like Lombardy[11], Emilia-Romagna[12], Piedmont[13] or Umbria[14]; by other territorial entities (like the Province of Trento Historical Archives[15]; by individual cultural or political institutions (like the Piedmont Institute for History of the Resistance and the "Giorgio Agosti" Contemporary Society[16], the Giangiacomo Feltrinelli Foundation[17], The Senate of the Repub-

[2] See its first version at <http://www.maas.ccr.it/h3/h3.exe/aguida/findex> and the new one, published online in 2009 at <http://guidagenerale.maas.ccr.it/>.
[3] See at <http://www.archivi-sias.it/>.
[4] See at <http://siusa.archivi.beniculturali.it/>.
[5] See at <http://www.archividelmediterraneo.org/portal/faces/public/guest/>.
[6] See the Florence State Archives Information System o rSiASFI at <http://www.archiviodistato.firenze.it/siasfi/>.
[7] See the Online Guide to the State Archives of Milan at , <http://archiviodistatomilano.it/patrimonio/guida-on-line/>.
[8] See the Archival heritage of the State Archives of Bologna at <http://patrimonio.archiviodistatobologna.it/asbo-xdams/>.
[9] See the Archival heritage of the State Archives of Naples at <http://patrimonio.archiviodistatonapoli.it/xdams-asna/>
[10] See the Online Guide to the State Archives of Venice or SiASVe at <http://www.archiviodistatovenezia.it/siasve/cgi-bin/pagina.pl>.
[11] See the section dedicated to the description of archives on the cultural heritage portal of Lombardy Region at <http://www.lombardiabeniculturali.it/archivi/>
[12] See the Ibc Archivies Portal at <http://archivi.ibc.regione.emilia-romagna.it/ibc-cms/>
[13] See Guarini Web for archives at <http://www.regione.piemonte.it/guaw/MenuAction.do>
[14] See .DOC- information System for the archives of the Umbria Region at <http://www.piau.regioneumbria.eu/default.aspx>
[15] See the online inventories on the web site of the Historical Archives of Trento Province at <http://www.trentinocultura.net/catalogo/cat_fondi_arch/cat_inventari_h.asp>
[16] See ArchOS. Integrated Siytem for the archival Catalogs at <http://metarchivi.istoreto.it/>
[17] See Foundation online archives at <http://www.fondazionefeltrinelli.it/feltrinelli-cms/cms.find?flagfind=quickAccess&type=1&munu_str=0_6_0&numDoc=95>

lic[18], The Chamber of Deputies[19] and many others) or by groups of 'federated' cultural institutes (the Institute of the Resistance Network[20] or the project Archives of the 1900s[21]).
The presence on the Web of these multiple systems is not only the result of the way in which historical archives have been computerized in Italy, nor is it just the result of the archival polycentrism mentioned above. Actually the dialectic between 'local' and 'national' systems of archival descriptions reflects a deeper logic that has to do with the double meaning associated with archives today. On one hand archives are products of specific historical and geographical contexts and hence, sources for knowledge of their history and vehicles of specific memories and identities; on the other, they are tout court cultural heritage, thus carriers of universal significance and values that cannot be closed within restricted territorial areas, but must acquire a national, and possibly international, visibility. Therefore, not only is this multiplicity of systems not casual, nor does it constitute a sort of limits to quickly overcome, possibly through their centralization into one single system, but contrarily, it represents an undoubtedly valuable resource. Nevertheless, it is true that in the previous years there has been a growing need to establish connections, data exchanges, increasing levels of interoperability between local, regional and national archival systems so that a great deal of reflections and discussions has taken place on how to create fruitful forms of collaboration[22].
By developing integrated access tools and offering essential information on the nation's archival heritage, CAT wants to build an answer to such a need and represent a tool for joining many existing systems together without substituting them; on the contrary it gives them greater visibility and enhances their specific characteristics. An operation like this is possible - even with the diversity of the software tools used and of some aspects of descriptive formats - since the systems developed in recent years share the same conceptual model and a common adoption of the international standards of archival description These systems are generally based on an architecture that, besides the descriptions of archival materials created in conformity with ISAD (G), include the separate description of creators (corporate bodies, families, persons) in accordance with ISAAR (CPF) and the description of custodians of archives (archival institutions, but sometimes other entities or families and persons as well), according to the International Standard for the Description of Institutions with Archival Holdings (ISDIAH).
Therefore, the CAT will sketch a general map of the national archival heritage capable of providing an initial orientation to researchers and guide them towards more informative resources available in the systems that will participate in the National Archives Portal. It will contain descriptive records of custodians, of fonds or archival aggregations, finding aids and creators. It will be populated and updated through procedures that privilege procedures of data harvesting based on the OAI-PMH protocol. Other ways of importing will not be excluded, as for example the upload of XML files in a specific area of the Portal or the direct entering data into the CAT database, using an ad-hoc on-line interface. The purpose of these multiple implementation procedures is to let even the less technologically equipped holder of archives participate in the project. .
Custodians of archival materials will be univocally identified and essential information on each of them will be provided in order to produce an authority list of all entities, institutional and non-institutional, that in Italy hold and provide access to historical archives. In addition to the essential identification data of each custodian (name, location) and a brief description of its history and current status, the availability of a reading room with regular opening hours and a reference service provided by skilled archivists will be indicated. Description and identification data, if available, will be acquired directly by the systems that participate in the PAN. The portal's editorial staff will correct, update and standardize the descriptions, if necessary, in order to provide users with reliable and up-to-date information. The CAT record for each custodian will contain, along with a hypertestual link to its web site, links to other descriptions which can be found in any of the systems participating in the PAN.

[18] See the Online Archives Project of the Senate of the Republic which includes descriptions and digital reproductions of archival fonds held also by other institutions at <http://www.archivionline.senato.it/>
[19] See web site of the Historical Archives of The Chamber of Deputies at <http://archivio.camera.it/archivio/public/home.jsp?&f=10371>
[20] See the Guide to Historical Archives of the Resistance Institutes at <http://beniculturali.ilc.cnr.it/insmli/guida.HTM>.
[21] See Archives of 1900 s - Memopry aon the Net at <http://www.archividelnovecento.it/archivinovecento/>.
[22] See for the example Verso un Sistema Archivistico Nazionale? special issue of "Archivi e Computer", XIII (2004), 2, edited by Stefano Vitali.

Regarding archival aggregations, the CAT will include the highest levels of description of each archive (fonds, or groups of fonds). It will also include lower levels of description such as sub-fonds or even series if created by specific creators, different from those of the fonds which they belong to.
The CAT will provide also concise information on the availability and main features of finding aids, existing on paper or in digital formats, for the archival aggregations described. An appropriate hypertextual link will address users to the digital ones directly accessible on the Web.
The selection of the information elements to be included in the CAT database for describing archival aggregation, creators, finding aids was prompted by a sort of principle of subsidiarity and economy and aimed to identify just those elements which are really essential. For wider and deeper descriptions, users will be addressed to the harvested systems. Consequently, only the elements considered mandatory according to the international descriptive standards together with few others regarded as such in Italian archival tradition have been included. A maximum number of characters in open-text fields will be provided in order to avoid redundant information. The description imported into CAT database from the provider systems will be published without any correction or modification.
Each CAT record will contain a direct link to the corresponding record in the original system, from which the data has been imported. Following the link the user will access the complete description of the archival aggregation, creator or the whole finding aid – if it exists - in its original context. The descriptions of archival aggregations will also be linked to relevant digital reproduction projects which will be made available on the Portal.
Each imported record is to be connected to a CAT record describing the original system in order to provide information about the provenance of the data. Moreover, since it cannot be excluded that the same archival aggregation, finding aid or creator is described in more than one of the archival systems exporting their data into the CAT database, provision of information about original systems will help users to correctly interpret and contextualize such multiple descriptions of the same entity.
Nevertheless, in oder to provide users with information of increasing quality and for adding value to records imported into the CAT database, the descriptions of creators will be linked to an authority file of creators which will be progressively implemented by the central staff of the Portal. This authority file should be not only the primary access point for research and navigation in the CAT, as well as the tool for connecting records imported from the provider systems, but also a national point of reference for identifying corporate bodies, persons and families and recording their authority names and descriptions. Therefore, in the future, local systems will not have to produce their own description of creators, but, if they would like, they could directly refer to the national authority file for creators published on the Portal, Finally, this authority file will represent a bridge towards analogous authority files developed in catalogs and descriptive systems of other cultural heritage domain, such as the National Library System.

SAN standards definitions and characters

The CAT architecture, the descriptive elements to be included and the exchange formats between the CAT itself and the systems which will provide the data have been developed by some ad hoc workgroups appointed at the beginning of 2009 by the State-Regions Joint Technical Commissions for defining archival standards. Representatives from different archival institutions and administrations and from various regions took part in the discussions. Besides metadata sets for the various entities described in CAT[23], the workgroups defined export-import formats, protocols and procedures and developed standards and methodologies for the preparation of the descriptions of corporate bodies, persons and families to be included in the authority file of creators. During the first few months of 2010 metadata sets will also be released for digital resources which will be made available on PAN, according to the same approach used in CAT. In the Portal's digital archives, thumbnails of the images and essential information, including hypertextual links to the original systems, will be imported, allowing users to research among the available digital resources, make a preliminary selection and then be directly addressed to the harvested systems for quality viewing of the digital reproductions of archival documents.

[23] See Sottocommissione tecnica per la definizione dei metadati […], Tracciati descrittivi del CAT: soggetti, <http://ims.dei.unipd.it/data/san/metadati/docs/2009-04-17_Documento-conclusivo-sui-tracciati-CAT.pdf> (provisional url).

For exporting the descriptions from the existing systems to the CAT, an XML exchange format has been developed. It has been named "SAN exchange format". It is based on three schemas, each of one includes a subset of elements of the Encoded Archival Description or of the recently released Encoded Archival Context (Corporate Bodies, Persons, Families)[24].
In particular, the schema for archival aggregations is based on EAD and include only elements from the <did> element. Due to some specific requirements of the SAN exchange format, many of the EAD elements used had to be partially adapted and modified. The schema for finding aids is based on EAD and includes only elements from the <header> element. Also in this case some of the elements have been adapted for meeting the specific requirements of the SAN exchange format. The schema for creators is based on EAC-CPF. In this case, the standard worked much better and no specific adaptation was required to develop a schema completely compliant with the characteristics of the SAN exchange format. Finally, a fourth ad hoc schema has also been developed for importing some essential information on custodians from the systems which can provide such information. As mentioned above these descriptions, after been imported, will be amended, standardized and updated by the editorial staff of the Portal[25].
In developing the schemas, an approach was adopted that can be defined as "recordcentric". Every item description exported will generate just one XML record, uniquely identified by the identity code of the description in the original system. Relationships between items - even those of hierarchical nature between archival aggregations – will be explicitly recorded, making reference to the identity code of the description/s of the related item/s.
The XML schemas for the four entities (custodians, creators, archival aggregations, finding aids), integrated with control information necessary for correctly implementing the import procedures, have been accommodated into a whole export-import format, which is available to all those responsible for the existing archive descriptive systems who wish export their data into the National Archival Portal and contribute to the effort in constructing a single access point to Italian archival resources on the Web[26].

Conclusions: Why the PAN experience is important

The realization of the National Archival Portal and its Catalog of archival resources is destined to represent an important turning point in the ways with which the Italian archival community has until now conceived and constructed its own presence on the Web. For the first time the community will have at its disposal a system built from the ground up through data exported from various providers. A system that wants to be completely placed on the inside of the Web's new horizons based on cooperation, interoperability and data reusability.
The Portal is going to change radically the Italian archival landscape, with significant impacts on the existing systems. The need to adapt, even if slightly, their data to make them fully compliant with the SAN exchange format will have inevitable fallbacks on the same systems that will be stimulated towards more elevated levels of homogeneity and standardization in order to allow wider and wider data reuse in contexts different from the original.
On the other hand, the open and cooperative model adopted for developing the SAN standards has already fulfilled largely satisfying and important results. Further consolidation of these results, which must derive from their wide use, will constitute an important premise for taking further steps forward in developing other more complex national exchange formats related to other aspects and components of archival description.
It is worth remembering that if standards and metadata are important because they make it possible for systems to exchange data, they are even more important because they force people and institutions to communicate.

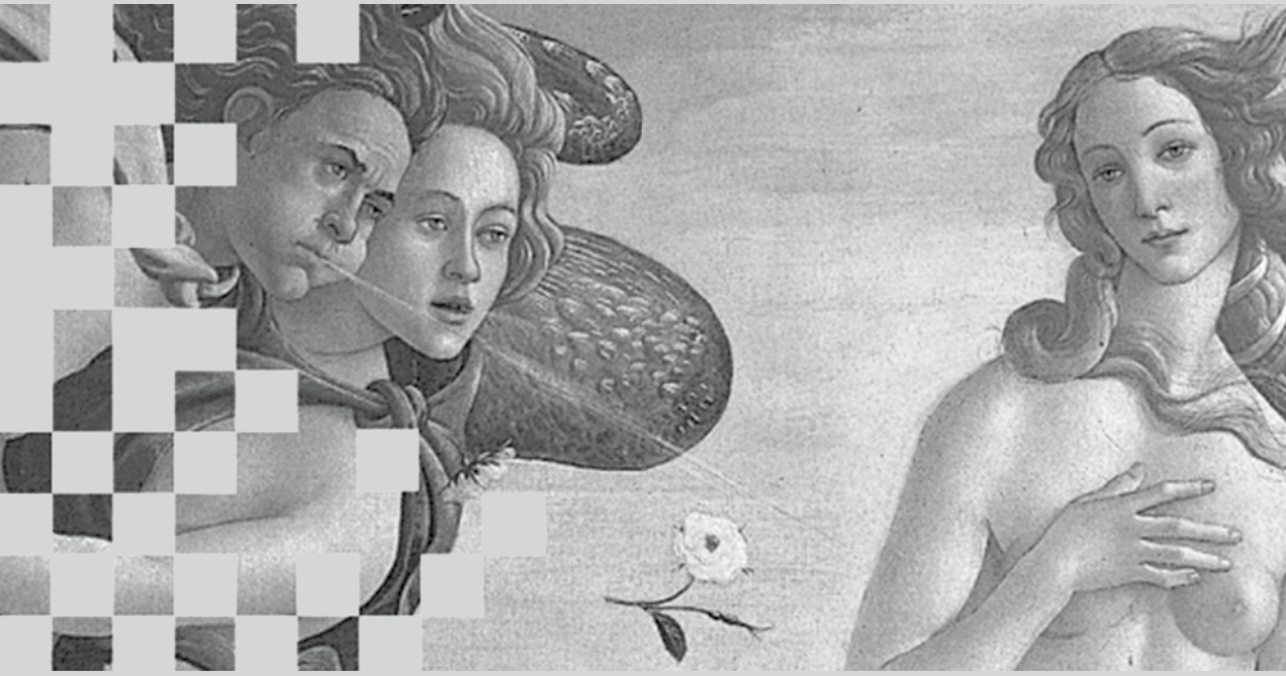
[24] See relevant documentation on the web site of the project at <http://eac.staatsbibliothek-berlin.de/>

[25] See Sottocommissione tecnica per la definizione dei metadati […],Tracciati descrittivi e schema XML di esportazione-importazione dai sistemi aderenti al catalogo delle risorse archivistiche (CAT) Versione 1.1, (ottobre 2009), <http://ims.dei.unipd.it/data/san/metadati/docs/2009-11-08_tracciati_export_import_1-1.doc> (indirizzo provvisorio).

[26] The overall schema and relevant documentation are provisionally accessible respectively at <http://gilgamesh.unipv.it/cat-import/cat-import.xsd> and <http://gilgamesh.unipv.it/cat-import/cat-import.html - id6>.

The Tuscan Region policies for the spread of digital culture go further on the specific policies about cultural heritage and I think it's right talking about them, because are critical for creating the environmental conditions to develop digital technologies for cultural heritage.
This work has begun with the establishment of a network, which function was regulated according to the regional law num. 1, 2004: "Disciplina della Rete telematica regionale Toscana". In the October 2009 a second regional law has been approved, it is the number 54: "Istituzione del sistema informativo e del sistema statistico regionale". In addition to this other intervention programs have been put in place for developing the implementation of digital technologies in local administrations, in local health authorities and in the justice management.
Regarding the specific sector of cultural heritage, we have to mention the important role that Tuscan public Libraries are carrying out in the promotion of digital culture during the last ten years, with the construction of new Libraries, some of them beautiful, and with the growth of services that implement largely digital technologies for users community.
The new Tuscan Libraries are doing a very important work to promote the digital culture and the rights for all to benefit from these technologies, information and their services.
About cultural heritage sector, the Region and his partners have promoted and supported several projects: some of them are already operative, e.g. The Digital Newspaper Library, others are still in execution, like the project for the digitization of papery catalogues of historical local Archives. We have published on line significant databases such as the D.B. of the ancient book, the mediaeval manuscripts, the Tuscan Lexicon Atlas and so on.
We have recently presented the DANTHE portal, created in collaboration with the University of Florence. This portal allows to access to the on line data bases regarding the cultural heritage in Tuscany.
We looked for and found a constant cooperation with the Ministry for Cultural Heritage and Activities and its structures in Tuscany. The results of this collaboration are: the beginning of the regional pole SBN, the participation in the MICHAEL project and in other projects related to Archives (SIUSA), the "Carta dei vincoli", and an on line data base about all the architectural, archaeological and landscape constraints.
The cooperation and coordination of the projects are of primary importance and represent the first target to be assigned to public administrations and, at the same time, are necessary for the availability of suitable financial resources.
Finally it's very important the integration between public institutions that manage cultural projects, from the State to local administrations, and private subjects like the Fondazione Rinascimento Digitale, that today has organized this Conference, but also other cultural institutions, e.g. universities, research Centres and enterprises.

Chiara Cirinnà and Maurizio Lunghi (edited by), *Cultural Heritage on line: Empowering users: an active role for user communities. Florence, 15th - 16th December 2009* ISBN 978-88-6453-184-7 (print) ISBN 978-88-6453-187-8 (online)



Parallel sessions I

Digital library applications & interactive Web

Introducing this session during the Conference Cultural Heritage Online I was looking for finding a reply to the following questions: How librarians and technologists can exploit the potential of emerging technologies? How can the cultural institutions transform the way digital libraries provide services and resources?
In trying to reply to these questions, after the conclusion of the session, I reflected on what the speakers have presented. They have described how emerging technologies can support the integration of different digital collections, can facilitate community building and extend connectivity to the ubiquitous user. However the speakers of the session have evidenced that there are challenges on existing delivery models of traditional cultural institutions, which have to change.
Two new roles were emerging from the presentations: a new role of digital libraries, a new role of users This two roles are two faces of the same coin: digital libraries are participating to a diffused culture of learning; users are actively engaged in creation, modification, and distribution of information objects in digital libraries as learners in a new virtual space.
Analyzing the different presentations, I can say that the digital libraries applications, with a focus on their users, move from the paradigm of cultural institutions as place to the paradigm of digital libraries as virtual spaces for learning.

New digital library role

Most of the speakers said that digital libraries should transcend the "search and access" approach, and serve as collaborative knowledge environments. The metaphor of the virtual space is very simple in technical term. The focus of the new role however is not on the tools and technologies, but on the changing roles of libraries supporting learning services. For example, the VKS (Virtual Knowledge Studio for the Humanities and Social Sciences) in Netherlands, focuses on the new scholarly practices, on the research methods, and on the ways of knowledge creation in the humanities and social sciences, stemming from the introduction of information and communication technologies into those scholarly fields.
Web 2.0 technology is a hot topic at the moment, and public librarians in particular are beginning to feel the pressure to apply these tools. Indeed, Web 2.0 has the potential to transform library services, but all the speakers have demonstrated that only if the policy and strategy for digital library services are ready to be transformed, the digital libraries can afford the new role. In a world where computing power, ubiquity and connectivity are increasing, the digital libraries could follow powerful new visions of their services for facilitating learning. This new digital library role brings cultural institutions into the cultural and social aspect of the technology.
Manzuch (Vilnius University) explained that the new role can be expressed as "communication of memory and involvement of the communities". This means bridging cultural heritage to the users, communicating heritages using the technology and digital media, sharing memory as part of the community values and cultural conditions. The present services of cultural institutions how are interpreting the cultural heritage? And how the services of cultural institutions are communicating memory?
With the contribution from leading practitioners in all areas – including lecturers, librarians and e-learning technologists – the session has explored the strategic approaches which digital libraries in the world are following. New services are the new access channels and value added services which EuropeanaConnect is going to provide for users of Europeana; or the service of digitisation on demand of the project EOD; or the activities supporting publishing of the EUI Library. "Interaction with the users" is the most advanced functionalities of the new services: transactions are made online, the whole ordering process and service management process (from the point of view of the digital library) is online and also modern payment services such as credit card payment – where needed are integrated in the services.
The speakers of the session have reviewed the technological tools of the digital library setting, but they have especially

Chiara Cirinnà and Maurizio Lunghi (edited by), *Cultural Heritage on line: Empowering users: an active role for user communities. Florence, 15th - 16th December 2009* ISBN 978-88-6453-184-7 (print) ISBN 978-88-6453-187-8 (online)

recommended the institutional policies, the theoretical reflections and the business models that are needed to create a new strategy for digital libraries. Challenges encountered when designing a large-scale application of digital library for a diverse public audience have been explored, including digital rights management, user content moderation, and balancing customizability with simplicity of interface. Also, staff development needs have been stressed, which includes programmes of training courses for the staff of the cultural heritage sector as well as higher and further education, conferences and workshop on best practices for exploiting the potential of the Web.

NEW USERS ROLE

How users can be involved in digital libraries? The first attempts made from digital libraries to include the active role of users in services have been described in the session: building networks of resources also produced by users, involvement of users in the selection of eBooks, participating to control the vocabulary, or putting data in the Web, collaborating in building repository of harvested resources. Users have been asked of leaving comments, annotating, making collaborative discussions, producing video content streams, cooperating with graphically enrich metadata visualisation. For example, the British Library (BL) is involving users in digitisation and content creation. BL asks to users text correction, improving the quality of the text version, putting their annotations, delivering user-created guides, participating to the discussion board and to the collaborative creation of text extraction.

More expert communities of users are involved in discovering, assessing, organizing critically all the primary resources, or in doing critical editions, not only terminological distinctions, putting annotations and doing linguistic analysis with syntactical tools, realising domain ontology to mark the texts of the documents.

The speakers have evidenced that we do not need pharaoinic projects but small scale projects, adapt to the moment and which fit in the different needs of the communities.

Some issues about the learner and teaching and learning interventions of digital libraries were also evidenced. Considering the enduring socio-psychological features of online interaction and the different user engagement, the diversity of user's community needs an interface design which is easy to access. User expectations want the accessibility of interface which should be focused on users intended purposes and the cultural institutions should not miss this opportunity of improving interaction and communication in the contemporary net-based social networks.

The most fundamental need in user's involvement is "trust": who is this person? The presentations gave a lot of information about user's generated content. Trust of communication and the expectations of an active role of users need a collaborative knowledge environment, as virtual space improving and facilitating learning.

CONCLUSION

Cultural institutions are going back to their roots, to start from their foundations to create new delivery mode of their services. The convergence in the Web of cultural institutions is not enough, if they offer services with no difference with the previous "search and access" paradigm.

Interistitutional cooperation needs to improve communication of memory and to create a virtual learning space. The focus on learning communities, scholars and other users move the cultural institutions services to the Web, moving from being a physical place to realising a virtual space, with the aim of promoting knowledge in extended services paradigms.

In conclusion the need of user's involvement is not new, the available tools are new and the functionalities can be highly extended, behind the traditional services. The projects in this session were able to use these new tools with creativity and the result was innovation.

Abstract
Following the initial excitement generated by Web 2.0, we are now seeing Web 2.0 concepts being adopted across the cultural heritage sector. Libraries, with their responsibilities for facilitating access to information resources and engaging with their user communities, have been early adopters of Web 2.0, and the term "Library 2.0" is now becoming accepted. Similar approaches are happening in the museums and archives sectors, with the terms "Museum 2.0" and "Archives 2.0" gaining currency.
But how should we ensure that the initial enthusiasms for use of Web 2.0 services and approaches become embedded within the organisation? And are cultural heritage organisations aware of the potential risks associated with making use of externally-provided services such as Facebook, YouTube and del.icio.us, including misuse of such services, associated legal concerns as well as the dangers of making use of services for which there may be no formal contractual agreements?
In this paper the authors argue that the cultural heritage sector needs to recognise that use of Web 2.0 providers does not necessarily provide an environment in which safe, secure and reliable delivery of services to the user community can be guaranteed. But rather than seeking to replicate successful Web 2.0 services in-house, we feel that we are in an environment in which cultural heritage organisations need to take a risk management approach to the use of networked services.
The paper describes a framework which is being developed, which aims to ensure that institutions have considered the risks associated with use of Web 2.0 technologies and services and have identified strategies for dealing with potential risks in order to achieve the goal of balancing the risks and benefits in order to maximise the dividends to be gained by use of Web 2.0.
Keywords: Social Web, Web 2.0, risks

Introduction

The Web 2.0 term has now been widely accepted as a description of a new pattern of ways in which the Web is being used. The Web environment has progressed from the publishing paradigm which characterised what is now sometimes referred to as Web 1.0, in which small numbers of content creators use tools ranging from desktop HTML authoring tools though to enterprise Content Management Systems and corresponding editorial and quality assurance processes to produce content for passive consumption by end users.
In a Web 2.0 environment large numbers of users are creating content using an ever-increasing variety of tools with such content being made available via a wide variety of commercial Web 2.0 services including photographic sharing services such as Flickr, video sharing services such as YouTube and social networking services such as MySpace and Facebook.
The characteristics of Web 2.0 were described by O'Reilly [1]. The key areas relevant to this paper include: (a) application areas including blogs and wikis, social sharing services and social networking services; (b) the ease of reuse of content elsewhere through syndication technologies such as RSS; (c) a culture of openness and sharing, which has been helped through the development of copyright licences such as Creative Commons; and (d) the concept of the 'network as the platform' by which services are hosted on externally-hosted services and accessible over the network, rather than a managed service within the organisation.
The Social Web is closely linked to Web 2.0. But whereas Web 2.0 includes various technical aspects (including technologies such as RSS and AJAX) the focus of the Social Web is very much focussed on the connections between people.

Opportunities

Why should cultural heritage organisations be interested in Web 2.0 and the Social Web? Answers to this question may include:
The Social Web is popular.

Chiara Cirinnà and Maurizio Lunghi (edited by), *Cultural Heritage on line: Empowering users: an active role for user communities. Florence, 15th - 16th December 2009* ISBN 978-88-6453-184-7 (print) ISBN 978-88-6453-187-8 (online)

- Social Web services can provide an opportunity to engage with new user communities and address challenges such as widening participation and social inclusion.
- Cultural heritage organisations, which are concerned with sharing and maximising access to cultural and scholarly resources, can exploit the Social Web to further this key mission.
- The Social Web can be cost-effective, allowing cultural heritage organisations to exploit a technical infrastructure that is already in place and is popular with many users.
- The economic downturn means funding for in-house development work is difficult to obtain.
- Popular Web 2.0 services can be easily used by end users as they can make use of services and interfaces they may already be familiar with.

These opportunities have been identified by many cultural heritage organisations which are already exploiting the Social Web's potential. Some examples of how museums, libraries and galleries are exploiting the Social Web are given below:
The National Library of Wales (NLW) has a remit to collect, preserve and give access to all kinds and forms of recorded knowledge, especially relating to Wales and the other Celtic countries, for the benefit of the public, including those engaged in research and learning. The use of Web 2.0 approaches for Library 2.0 delivery is ingrained in the NLW's 2008 strategy document Shaping the Future [2] which outlines the Library's desire to explore collaborative and diverse models using external resources. This will allow the NLW to leverage Web platforms which are heavily focused on user engagement in order to deliver future services.
The Brighton on the Pull Project provided an opportunity for Brighton Museum & Art Gallery to work with target audiences and new ways of researching their collections. The ethos of On the Pull was concerned with taking a step away from the traditional museum exhibition to encourage new visitors and target audiences. The project team explored use of social networking services as a marketing tool in order to get away from the associations with the word 'museum' as a way of breaking down barriers and the connotations the word was found to hold for the focus groups. Music and video clips hosted on YouTube are embedded in MySpace. In addition to MySpace, FaceBook was also used as a marketing tool to advertise events, promote competitions, display promotional images, images of objects from the collections and play music. [3].
The arrival of a pair of nesting peregrine falcons at Derby Cathedral provided Derby Museums with an opportunity to promote the town to a large audience. A Webcam provided live video footage of the nesting of three chicks and an accompanying blog and MySpace account, together with use of Flickr and YouTube for providing access to photographs and video footage resulted in "evidence emerging of visitors coming to Derby specifically because of its peregrines" [4].

Risks and Barriers

Identifying the Concerns

UKOLN, a national centre of expertise in digital information management based at the University of Bath has, over the past two years, delivered a series of workshops for the UK's cultural heritage sector. The workshops have provided an opportunity for practitioners in the sector to gain an understanding of the potential of Web 2.0 and to explore its potential. The workshops have also identified barriers to the effective deployment and use of the Social Web. A summary of the various concerns is given below.

Sustainability Challenges: There may be concerns over the lack of interoperability of third party services, with dangers that a service may be a 'walled garden', allowing data and content to be added to the service or created within the service but cannot be exported to another environment.
Technical Challenges: IT support staff may raise technical concerns related to reliance on third party organisations to deliver services for the organisation. These concerns might include performance and reliability issues, security, backups, etc.
Interoperability Challenges: Technical concerns raised may also cover the interoperability of third party services with other systems. This might include integration with existing in-house services and the export and migration of data to other services, including replacement services which might not be currently available.

Support Issues: Although many popular Social Web services can be used without formal training or support, use of the services in an institutional context may generate user queries.
Individual Concerns: Individuals within the organisation may be concerned with the deployment of Social Web services. Staff within the organisation may be reluctant to use technologies such as blogs and micro-blogging services such as Twitter because of an unfamiliarity with the technologies or the culture and expectations in these technologies or a desire to keep professional and social activities separate.
Organisational Issues: Proposals to make use of Social Web services by a cultural heritage organisation may not be universally welcomed by everyone within the institution. This may be regarded as undermining the organisation or a department in the organisation. Such concerns may not be openly articulated, but may lie behind concerns raised listed above.

The Legal and Related Concerns
There are a number of legal risks involved in creating and using resources hosted on Social Web services. Briefly, they can be summed up as follows:
Putting materials in that one should not deposit, because the copyright or other Intellectual Property Rights are held by third parties.
Use of Registered Trade Marks or unregistered trade names without permission.
Failure to identify someone as an author when they should be so named – this may well be an infringement of their Moral Right of paternity.
Failure to respect the Moral Rights of authors e.g. derogatory treatment of their work.
Data or advice that is inaccurate or misleading, and could lead to financial, physical or other damage to third parties if followed.
Outputs that break the Data Protection Act or infringe personal privacy.
Outputs that contain illegal materials e.g. materials that are pornographic, encourage terrorism, are defamatory, are in Contempt of Court, break race or sex discrimination laws, etc.
Outputs that break the Disability Discrimination Act by being unreadable to those with impairments.

A Risks and Opportunities Framework

The Tensions
This paper has provided examples of use of the Social Web by cultural heritage organisations which have identified the benefits which the Social Web can provide in enhancing the range and quality of services to the organisations' user communities. However we have also listed a range of concerns which organisations considering making use of Social Web services will need to consider. We will now describe a risks and opportunities framework which has been developed in order to support cultural heritage organisations in making decisions on use of the Social Web.

A Risks and Opportunities Framework For Addressing The Tensions
A risks and opportunities framework has been developed to support cultural heritage organisations in making effective use of the Social Web [6]. This paper introduces further developments to the framework including a summary of risk minimisation approaches and an inclusion of an evidence base.
Use of the framework involves documenting the following aspects of the proposed use of the Social Web:
Intended use: Rather than talking about Social Web services in an abstract context ("shall we have a Facebook page" for example) specific details of the intended use should be provided.
Perceived benefits: A summary of the perceived benefits which use of the Social Web service are expected to provide should be documented.
Perceived risks: A summary of the perceived risks which use of the Social Web service may entail should be documented.
Missed opportunities: A summary of the missed opportunities and benefits which a failure to make use of the Social Web service should be documented.

Costs: A summary of the costs and other resource implications of use of the service should be documented.
Risk minimisation: Once the risks have been identified and discussed approaches to risk minimisation should be documented.
Evidence base: Evidence which back up the assertions made in use of the framework.
When using this framework it should be recognised that there are likely to be biases, prejudices, vested interests and other subjective factors which will affect how the framework is used. Ideally such subjective factors will be openly acknowledged and taken into account, although it is recognised that this may be difficult to achieve.

Application to In-House Developments and Existing Services
It should be noted that this framework need not only be applied to proposals to make use of the Social Web. In order to minimise the subjectivity of the approach it should also be applied to proposed in-house development work and commissioning IT developments. It can also be applied to existing services in order to identify the risks, limitations and constraints which the organisation is willing to tolerate and accept.

Legal Risks

A Risk Assessment Formula for Legal Infringements
The risks and opportunities framework recognises that although there will be risks when seeking to exploit the Social Web, it may be necessary to accept such risks in order to deliver services to the user community. A similar approach can be taken to addressing the risks associated with possible copyright infringement.
The example below relates to copyright infringement, but the same formula applies to all the legal risks identified above. The risk can be calculated as follows:
R = A x B x C x D
where R is the financial risk; A is the chances that what has been done is infringement; B is the chances that the copyright owner becomes aware of such infringement; C is the chances that having become aware, the owner sues and D is the financial cost (damages, legal fees, opportunity costs in defending the action, plus loss of reputation) for such a legal action. Each one of these other than D ranges from 0 (no risk at all) to 1 (100% certain). D is potentially a high number. It is not easy to calculate the cost of loss of reputation.
Factors to bear in mind:

- If the work is to be used in a commercial context (i.e. to generate financial gain) then a rights owner who later becomes aware of the use of their work may be more likely to pursue an action for infringement of copyright than if the work is being purely used for educational purposes.
- The nature of the content used, for example, the rights in high value content, such as commercially produced films, text, images, music and software, are more likely to be actively enforced by their owners.
- Particularly sensitive subject areas are music, geographic data, literary works by eminent authors, and artistic works including photographs and drawings.
- Is there any track record of the contributor ignoring legal niceties in the past?
- Is there any track record of a particular third party having complained before?

Depending on these factors, the risks will vary. However, a Web 2.0 provider that ignored warning signals (e.g. a contributor who ignored legal niceties in the past is allowed to continue to add more materials without checks being made) is likely to receive an unsympathetic hearing from the Courts. Similarly, a service provider who has failed to educate contributors regarding legal issues will also not be viewed sympathetically by a Court.
Ultimately, it is important that the service provider is proportionate about possible risks whilst at the same time prepares suitable mitigating strategies in the eventuality of a complaint. An apology and promise of swift action to rectify is often sufficient.

Reducing the Legal Risks

A cultural heritage organisation making use of the Social Web should ensure that it has clear and robust notice and take down policies and procedures with a clear address given for complaints. The notice and take down policy adopted by JORUM would be a valuable starting point [7]. Clear instructions should be given as to where and to whom notification of allegedly illegal content should be sent, along with details of the complainer, the complainer's interest in the matter and where the complainer can be contacted. Processes should be put in to place to act expeditiously on such a notification.

Conclusions

We have seen how cultural heritage organisations are successfully exploiting the Social Web in order to deliver quality services to their user communities. There is, however, an awareness that there are a variety of risks associated with use of the Social Web. The use of a risks and opportunities framework and a risk assessment formula have been described which aim to support the discussions and policy-making processes organisations need to take when formulating policies on exploiting the Social Web.

References

[1] What Is Web 2.0? O'Reilly, T. Retrieved September 20th 2009: http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html
[2] Shaping the future: The Library's strategy 2008-2009 to 2010-2011, National Library of Wales. Retrieved September 20th 2009: http://www.llgc.org.uk/fileadmin/documents/pdf/Strategy2008-2011.pdf
[3] On The Pull, English, C. UKOLN Cultural Heritage blog, 2 April 2009. Retrieved September 20th 2009: http://blogs.ukoln.ac.uk/cultural-heritage/2009/04/02/the-on-the-pull-project/
[4] When Peregrines Come To Town, Moyes, N, UKOLN Cultural Heritage blog, 18 May 2009. Retrieved September 20th 2009: http://blogs.ukoln.ac.uk/cultural-heritage/2009/05/18/when-peregrines-come-to-town/
[5] Time To Stop Doing and Start Thinking: A Framework For Exploiting Web 2.0 Service, Kelly, B., Museums and the Web 2009 Conference. Retrieved September 20th 2009: http://www.ukoln.ac.uk/web-focus/papers/mw-2009/
[6] JORUM procedures to Deal with Queries, Alerts and Complaints, JORUM. Retrieved September 20th 2009: http://www.jorum.ac.uk/docs/pdf/takedown_procedure.pdf

Abstract
This paper is part of the authors' joint project on trust in online interaction, and it contributes to the enhancement of collaborative knowledge environments by advancing our understanding of key socio-psychological features of online communities and user engagement. We first review the history of online communities and user engagement, focusing on the socio-psychological features of trust in online interaction. Next, we discuss the issue of trust with regard to user-generated content and cultural heritage, highlighting the issues of trusting beliefs, trusting intentions, and trust transfer. Finally, we argue that a diachronic understanding of online practices holds the capacity to explain much of what we see online today, and we propose that the power of this digital legacy should particularly be valued and employed in the institutions and contexts promoting and fostering cultural heritage in both traditional and contemporary forms.
Keywords: trust; user engagement; Internet; cultural heritage.

Introduction

The role of users and user communities in the production of online content has become a hot topic in recent academic and public debate, featuring themes such as social networking, user labor, user co-creation, user-generated content (UGC), social annotation, folksonomies, and the like. As van Dijck [1] points out, "with the emergence of Web 2.0 applications, most prominently UGC platforms, the qualification of 'user' gradually enters the common parlance" (p. 41). In this paper, we show that user engagement and user generated online content have a much longer tradition than the current debate implies, and we argue that a diachronic understanding of online practices holds the capacity to explain much of what we see online today. The idea of engaged users and engaged user communities was in a sense "hard wired" into the very earliest pre-Internet systems, thus providing the basis for what we have today and more importantly, for what we have come to expect and even demand in online systems.

Contemporary notions such as "social Web," "participatory Web," "prousers" and the like obscure the fact that the earliest computer networks had been built as networks of people, not wires, and have always been social and participatory, even before they became the Web. What Web 2.0 has brought is rather the matter of performance than of essence; it has brought easy to use content-creating applications - such as blogs, wikis, social networking sites, and file sharing platforms - rooted in broadband access, affordable hardware and software solutions, and with the Internet perceived and used as a "new normal" in contemporary way of life [2]. Put differently, Web 2.0 has made visible what has always been there - engaged users and engaged online communities.

Trust in Online Communities and User Engagement

The concepts of online communities and user engagement have always been central to the analyses of online interaction, even in the pre-Web days. For example, in their prescient 1968 paper, Licklider and Taylor [3] observed that the very earliest networked computers of their time enabled for formation of communities, noting, in a vision that has become the underlying theme of all research in online community to date, that "[online communities of the future] will be communities not of common location, but of common interest" (pp. 37-38; emphasis in the original). As networked computing moved from the laboratories into universities and the corporate world, researchers such as Hiltz and Turoff [4], Kiesler, Siegel and McGuire [5], and Turkle [6] continued to fine-tune our understandings of the social-psychological features of computer-mediated communication (CMC). And as networked computing moved from internal sites (such as corporations and universities) into the general populace, the study of online communities shifted to a wider range of cases and academic disciplines, providing both optimistic [e.g., 7;8] and pessimistic [e.g., 9;10]) interpretations of online sociality.

Gradually, new dynamics involving trust and credibility of online interaction began to emerge, highlighting the issue of user engagement as an inseparable element of online sociality. For instance, in her comparative study of two online protests, Gurak [11] demonstrated that already in the pre-Web days a large number of users could quickly assemble

around an idea of common interest, and, by sharing information and providing user generated content, create an efficient and successful online social action. Baym's [12] work on online soap opera fan clubs and Hine's [13] methodologically-oriented study of online activities surrounding the Louise Woodward case offered additional research-based evidence for the power of user-provided content and online communities, while Bakardjieva [14] showed that users transcended the sphere of personal experience by engaging in collective online practice; she asked why the users did what they did and what it mean to them. The same questions - why do users engage in providing online content and what does it mean to them - have come to the fore in the recent analyses of UGC, such as Baym and Burnett's [15] study of amateur experts' provided online content. This last point is one to which we will return in the next segment, when we discuss the main features of user-generated content, particularly in regard to cultural heritage.

In summary, contemporary online forums facilitate and promote features of online discourse that have been part of online interaction from the outset. The gathering of like-minded people around communities of common interest is key to understanding user engagement in the 21st century. Today's user wants quick, accurate, customizable, smart systems, and they want systems they can trust.

Indeed, all forms of human communication, but particularly online communication, depends heavily on trust. As Seligman [16] has stated, "[t]he existence of trust is an essential component of all enduring social relationships" (p. 13). Drawing upon the work of Luhmann, Seligman notes the relationship between trust and confidence, the latter based on whether "one can rely or place confidence in the other's words or commitments or acts" (p. 21; emphasis in the original). Trust, in these terms, then, "involves a vulnerability occasioned by some form of ignorance or basic uncertainty as to the other's motives," which Seligman notes as a particularly interesting concept in the Internet age, because of the "fundamental opaqueness toward the will of another" (p. 21). In other words, while trust and confidence in others is a foundational concept in all forms of human communication, it is a particularly interesting one in the digital age. We could not have functioning online communities without trust; indeed in broad terms, trust is "an important dimension of civic culture" [17, p. 14]. And we could not have any level of user engagement without both trust in the system and confidence in the motives of the system itself as well as the motives of other online participants.

Numerous studies and commentaries have been written on the issue of trust in digital environments in relation to issues such as usability [18], e-commerce [19], interface design [20], credibility [21], and other areas. Bailey, Gurak, & Konstan [22] have noted that "[t]rust plays a critical role when a user assesses the believability of online information content or when selecting an exchange site to purchase a product from. Users will not believe or participate in a transaction with those whom they do not trust" (pg. 311). Bailey, Gurak, & Konstan thus define trust as "the perception of the degree to which an exchange partner will fulfill their transactional obligations in situations characterized by risk or uncertainty" (p. 313), and they posit seven dimensions of trust in digital settings: attraction, dynamism, expertness, faith, intentions, localness, and reliability (p. 315).

More applicable to this paper, however, is recent work on trust in digital repositories [23]. Summarizing the literature regarding technical considerations when building a trusted digital repository, Prieto notes issues such as persistent access, content migration, resource discovery, data collection/quality, and so on (p. 596). He also notes the importance of users feeling that the content within a digital repository is itself trustworthy (p. 596). Yet he then makes this point: the role of the digital repository's stakeholders (whether they are referred to collectively as a community or individually as depositors or users) is key to establishing trust. Put another way, he says that "the repository can be trusted because it has been deemed an appropriate place into which content can be contributed or from which content can be retrieved for purposes of research, study, enrichment, or personal enjoyment" (p. 596). Prieto then goes on to describe the kinds of incentives that might need to be put in place in order to move the academic world from trust that is rooted in print to trust in digital repositories (p. 597).

To this end, we feel that user-generated content, a key feature of the Web 2.0 age, would build on the features of online community as well as digital trust, and could be of key importance to online systems designed for cultural heritage in the 21st century.

Trust in User-Generated Content and Cultural Heritage

User-generated content refers to online content produced by end-users. Online material is considered user-generated when/if it is publicly available, created outside of professional settings, and includes a user's creative effort (i.e., the

user did not simply copy and paste the content, but rather has added his or her own creative value to it); [24]. The practice of creating and providing user-generated content is usually considered to stem from the following motives: obtaining public acknowledgement; earning peer recognition; building reputation in a community; expressing oneself; developing skills that can become a profession; having fun; sharing knowledge/contributing to a common idea. In the case of UGC related to cultural heritage, sharing knowledge and contributing to a common idea is often seen as key stimuli, although other motives also play the role in this type of user engagement. For instance, both professional and amateur subject specialists are often prompted to contribute specialized local and/or minority content in different languages, and/or to engage in cultural heritage tribute online activities [25].

User-generated content that contributes to the cultural heritage sector is deemed to have both positive and negative aspects. For instance, UGC is often considered to complement and enhance institutionally provided content by offering novel information on specific—often local—cultural phenomena, as well as by offering novel ways of presenting and/or interpreting those phenomena. UGC is also regard as a means of transforming static content authority into dynamic, multisided knowledge platform, which has the capacity to engage the public as an acknowledged knowledge-provider [26, p. 20]. Finally, users' active online engagement with cultural phenomena is considered to have broader implications for the perception of cultural heritage, by fostering understanding of culture as an ongoing process, not as closed, historic experience completed and rooted in the past [25]. However, UGC is also considered to introduce unverified and/or difficult to verify popular knowledge (sometimes called "crowdsourcing") into the cultural heritage domain. Participatory culture prompted by Web 2.0 applications especially raises the question of trusted and reliable content with regard to sensitive multicultural issues and/or issues stirring intercultural debates. Similarly, UGC prompts the question of intellectual property (IP), either in the sense of granting users IP rights over their online creations, or in the sense of insuring that UGC acknowledges authorship of the original source(s) when needed.

Both positive and negative aspects of user-generated cultural heritage content bring us back to the subject of trust, by invoking the issues of trusting beliefs, trusting intentions, and trust transfer [27, p. 21]. While trusting beliefs refers to a person's perception of a certain actor and/or source as trustworthy, trusting intentions encompass a person's willingness to make him or herself vulnerable in accordance with such a perception. With UGC, trusting intentions encompass users' willingness to make themselves vulnerable to cultural information provided by other users. Vulnerability in this case implies exposure to potentially incorrect, incomplete, misleading, biased, or in some other way corrupted content. On such occasions, the issue of trust-transfer often comes to the fore, requiring that the burden of establishing trust be transferred from the user to an external proof source. In case of UGC, the proof source is usually institutional and/or contextual. For instance, if a piece of UGC is provided within the official website of an acknowledged cultural institution, the cognitive mechanism of trust-transfer associates such a piece with the given organization, reassigning the proof burden from the user to the institution. Similarly, if a particular user-generated post is provided within an acknowledged online resource, such as Wikipedia, and/or within an online community known to the post recipient, the trust-transfer process shifts the proof burden to the given online context. User-generated content thus reflects a complex interplay among users, institutions, and online contexts, or, as Ridge [28] points out, it reflects the issues of sharing authorship and authority with regard to cultural heritage. To address those issues, both the users and the cultural heritage institutions should be empowered to efficiently employ online systems designed for cultural heritage in the digital age. User empowerment in this sense implies providing customizable, smart, and easy to use systems, which enable users to employ their knowledge, creativity, interest and other positive stimuli by creating, sharing, and discussing cultural content. Similarly, institution empowerment implies enabling cultural heritage organizations to employ the power of user engagement, while at the same time avoiding previously mentioned downsides of UGC through sets of relevant procedures [28]. In other words, developing trust both within and among users and institutions is key to empowering these major actors in contemporary processes of cultural production and preservation.

Conclusion

Research on digital technology and culture has always made clear that the power and potential of the Internet lies in the unique dynamics and possibilities for online communities and user engagement. Nowhere is this concept of user and communal power more clearly visible than with today's social networking applications and Web 2.0 forms. Contemporary users expect to play an active role in shaping their online experience and want systems that are smart, customizable,

and cross traditional boundaries. Consequently, cultural heritage institutions should transcend the "search and access" approach, and in contract should serve as collaborative knowledge environments that promote users actively engaged in creation, modification, and distribution of information objects. To achieve this, cultural heritage institutions have the potential to employ valuable knowledge on user practices harvested in the field of Internet Studies over the past thirty years. While technologies and forms of online interaction have been and will keep changing in a blink of the eye, the main socio-psychological features and dynamics of user engagement have been and most likely will remain rather steady across the platforms and contexts of use. The power of this digital legacy should particularly be valued and employed in the institutions and contexts promoting and fostering cultural heritage in both of its traditional and contemporary forms.

References

[1] Dijck, van J. (2009). "Users like you? Theorizing agency in user-generated content." Media, Culture & Society, Vol. 31 (1), pp. 41-58.
[2] Pew Internet & American Life Project. (2005). Internet: The mainstreaming of online life. Retrieved June 27, 2007, from www.pewinternet.org/pdfs/Internet_Status_2005.pdf.
[3] Licklider, J. C. R. and Taylor, R. (1968). "The Computer as a Communication Device." Science and Technology, April: pp. 21-41.
[4] Hiltz, S. R., & Turoff, M. (1978). The Network Nation. Reading, MA: Addison-Wesley.
[5] Kiesler, S., Siegel, J., & McGuire, T. W. (1984). Social and psychological aspects of computermediated communication. American Psychologist, 39(10), pp. 1123–1134.
[6] Turkle, S. (1984). The second self: Computers and the human spirit. New York: Simon & Schuster.
[7] Rheingold, H. (1993). The virtual community: Homesteading on the electronic frontier. New York: Harper Collins.
[8] Reid, E.M. (1991). Electropolis: Communication and Community On Internet Relay Chat. Retrieved October 15, 2009, from http://irchelp.org/irchelp/misc/electropolis.html
[9] Doheny-Farina, S. (1996). The Wired Neighborhood. New Haven, CT: Yale University Press.
[10] Fenerback, J. and Thompson, B. (1995). Virtual Communities: Abort, Retry, Failure?. Retrieved October 15, 2009, from http://www.well.com/user/hlr/texts/VCcivil.html.
[11] Gurak, L. J. (1997). Persuasion and privacy in cyberspace: The online protests over Lotus market place and the clipper chip. New Haven, CT: Yale University Press.
[12] Baym, N. (1999). Tune in, log on: Soaps, fandom, and online community. Thousand Oaks, CA: Sage.
[13] Hine, C. (2000). Virtual Ethnography. London: Sage.
[14] Bakardjieva, M. (2003). "Virtual togetherness." Media, Culture, & Society, Vol. 25 (3), pp. 291-313
[15] Baym, N.K., and Burnett, R. (2009). "Amateur Experts: International fan labour in Swedish independent music." International Journal of Cultural Studies, Vol. 12(5), pp. 433-449.
[16] Seligman, A. B. (1997). The problem of trust. Princeton, NJ: Princeton University Press.
[17] Sztompka, P. (1999). Trust: a sociological theory. Cambridge, UK: Cambridge University Press.
[18] Bedi, P. and Banati, H. (2006). "Assessing user trust to improve web usability." Journal of Computer Science, Vol. 2 No. 3, pp. 283-7.
[19] Klang, M. (2001), "Who do you trust? Beyond encryption, secure e-business," Decision Support Systems, Vol. 31 No. 3, pp. 293-301.
[20] Pu, P. and Chen, L. (2007), "Trust-inspiring explanation interfaces for recommender systems," Knowledge-Based Systems, Vol. 20 No. 6, pp. 542-56
[21] Fogg, BJ & Tseng, H. (1999). "The elements of computer credibility". Proceedings of ACM Conference on Human Factors and Computing Systems, pp. 80 – 87.
[22] Bailey, B.P., Gurak, L. J., and Konstan, J. (1998). Trust in cyberspace. in Human Factors and Web Development. ed. Ratner, J. Hillsdale, NJ: Lawrence Erlbaum Associates, pp. 311-321
[23] Prieto, A. G. 2009 From conceptual to perceptual reality: trust n digital repositories. Library Review 58 (8), pp. 593-606.
[24] OECD. (2007). Participative Web and User-Created Content: Web 2.0, Wikies, and Social Networking. Retrieved October 15, 2009, from http://213.253.134.43/oecd/pdfs/browseit/9307031E.PDF.
[25] Harrison, R. (2009). "Excavating Second Life: Cyber-Archaeologies, Heritage and Virtual Communities." Journal of Material Culture, Vol. 14 (1), PP. 75-106.
[26] Coppola, P., Lomuscio R., Mizzaro, S., Nazzi, E., and Vessena, L. (2008). "Mobile Social Software for Cultural Heritage: A Reference Model." 2nd Workshop on Social Aspects of the Web (SAW 2008), BIS 2008 Workshop Proceedings, pp. 69-80.
[27] Stewart, K.J. (2003). "Trust Transfer on the World Wide Web." Organization Science, Vol, 14 (1), pp. 5-17.
[28] Ridge, M. (2007). Sharing authorship and authority: user generated content and the cultural heritage sector. 2007 Web Adept - UK Museums on the Web, Museums Computer Group conference. Retrieved October 15, 2009, from http://www.miaridge.com/projects/usergeneratedcontentinculturalheritagesector.html

Abstract
This paper provides an outline of the main goals of the project EuropeanaConnect, one of the two core technical projects in the Europeana project cluster which will develop the current prototype of Europeana into a fully operational service until 2011. Europeana is Europe's digital library, archive and museum portal, launched in November 2008.
Keywords: Europeana, semantic web, multilingual access, mobile access, value-added services, licensing framework, audio content

Introduction

The prototype of Europeana (http://www.europeana.eu), Europe's digital library, archive and museum portal, was launched in November 2008. The European Commission is funding the development of Europeana as part of its i2010 Digital Libraries initiative to create an integrated access channel to Europe's distributed digitised cultural heritage resources.

The launch of the Europeana prototype created such a large-scale public interest that the site had to be taken down for a few weeks for a substantial hardware and load balancing reconfiguration. Back online in a test mode since December 2008 and with full functionality since April 2009, the Europeana prototype currently gives access to about five million digitised objects of cultural value from over 1.000 European libraries, archives, museums and audio-visual collections. Within the next two years Europeana will be developed into a fully operational service. By mid 2010 Europeana is expected to provide access to 10 million items; the target for 2012 is 25 million items. The European Commission is funding a number of projects in the eContentplus and ICT PSP programmes. Within the next years, these projects will develop the technology and services for Europeana and digitise and aggregate content for the portal.

Two core projects will be responsible for the implementation of the technical infrastructure of Europeana: The project Europeana v1.0 (http://version1.europeana.eu/) is led by the European Digital Library Foundation. This 30-month project which started in February 2009 will solve key operation issues related to the implementation of Europeana like building the back end services needed to manage the delivery and access of Europeana content and managing the channels that will enable other environments to use the content made interoperable by Europeana via web services and APIs. This project will also develop and manage partnerships to ensure a rich content flow from national and domain aggregators. The second core project is EuropeanaConnect (http://www.europeanaconnect.eu) and will be described in the remainder of this paper.

EuropeanaConnect

EuropeanaConnect is closely associated with the project Europeana v. 1.0 and is a 30-month Best Practice Network co-funded by the European Commission in the eContentplus programme. Co-ordinated by the Austrian National Library, EuropeanaConnect will deliver core components and value-added services for Europeana. It will also act as a content aggregator for audio content and will significantly enlarge the music content available in Europeana. The project runs from May 2009 to October 2011 and brings together 30 partners from 14 European countries, including universities, research institutes, libraries, audio archives and a partner from the publishing industry.

Europeana Semantic Layer

One of the main goals of EuropeanaConnect is to build a layer of semantic data which will be the basis for all semantic processing in Europeana. Semantic processing would, for example, enable a relationship between "Mona Lisa", "Lisa del Giocondo", "La Joconde" and "Leonardo da Vinci" to be recognised during the search process. A user entering just one of these terms would see also results for the linked concepts, e.g. would automatically find other paintings created

Chiara Cirinnà and Maurizio Lunghi (edited by), *Cultural Heritage on line: Empowering users: an active role for user communities. Florence, 15th - 16th December 2009* ISBN 978-88-6453-184-7 (print) ISBN 978-88-6453-187-8 (online)

by Leonardo da Vinci. EuropeanaConnect will provide the technologies to semantically enrich huge amounts of digital content in Europeana. This will enable semantically-based content discovery and make Europeana content more accessible, reusable an exploitable.
EuropeanaConnect will create a network of semantic resources starting from controlled vocabularies and other similar resources which will be provided by content owners. These will be made ready for semantic processing mostly by using SKOS (Simple Knowledge Organization System, a W3C standard) which in turn will be used as the primary level of user interaction with Europeana. Unlike in traditional library catalogue-driven models of user interaction, users of Europeana will be able to explore the data space by using the semantic nodes (i. e. concepts organised and available as web resources) as primary elements for searching and (faceted) browsing.

Multilingual Access
Multilingual and multicultural aspects are at the heart of making Europeana's digital contents effective and exploitable for users across all European countries. Metadata and digital content should be searchable and presentable independent of the language of the searcher or the object descriptions. In other words, users should be able to enter search terms in their own or preferred language, which are translated on the fly during the search process, and have hits returned in a range of selected languages. In order to enhance the multilingual access capabilities of Europeana, EuropeanaConnect will develop a multilingual infrastructure and a set of translation tools that will process queries and object data and produce a suitable multilingual representation for the user:

- a unique repository of language and translation resources for multilingual processing of objects within Europeana, which will possibly be one of the most varied of its kind in Europe;
- a suite of tools for multilingual mapping of controlled vocabularies in at least five European languages and a strategy for incorporating more languages;
- a suite of integrated translation modules for querying and browsing processes which will plug into the general search and retrieval infrastructure of Europeana.

Not every European language is equally well developed in terms of multilingual resources for translation. EuropeanaConnect will start with a core set of languages (English, French, German, Italian, Spanish) for which expanded multilingual translation capabilities will be implemented by the end of the project. During the course of the project, the core set will be expanded, based on the development experiences, with a set of additional languages integrated with less linguistic treatment based on available resources and content.

Tools and methodologies for user-driven development
EuropeanaConnect will provide a better understanding of what current and future web-users really need and expect of the multi-lingual and multi-channel service which Europeana must become. The project will provide the methodology to monitor and evaluate the user interactions with Europeana. Deep Log Analysis will allow for the evaluation of significant information-seeking behaviour of Europeana users. This method will be used by Europeana as a routine for understanding how the Europeana portal is used and what its users really expect from it. The usage-monitoring methodologies and the results of Deep Log Analysis will inform the implementation iterations of the Europeana portal (Europeana v1.0) and will thus ensure that the development of Europeana is truly user-driven.
Europeana sees all groups of people as potential users. It will not be sufficient to have information on the actual user behaviour; we also need to move from the concept of the user as "anonymous" to an understanding of who is using what. To support this and to generally ensure user-oriented service development and marketing of Europeana, EuropeanaConnect will define "personas", i.e. categories with profiles. We need to understand the link between the results of the deep log analysis and these personas and we need to understand how Europeana can add value to a selected number of these personas. This work will be accompanied by actual user studies either through questionnaires or via different forms of usability tests. From a preliminary screening we aim for a rough number of 500 people to be involved throughout the lifetime of the project.

New Access Channels to Europeana

Spatio-temporal user interface
EuropeanaConnect will develop a spatio-temporal access service that will support access to Europeana via a number of space- and time-related selection mechanisms as part of the query- and result-presentation process. It will implement a user-friendly access mechanism, which exploits the Semantic Layer of Europeana. The objective is to create an interface that makes use of both time-based and geographical metadata of the Semantic Layer, providing a new visual access to the digital collections in Europeana. The project will build an interactive and generic map combined with a timeline, acting as a "histogram" for tagged events, places and characters in Europeana. The interactive map will show the visual representation of a search-result set in Europeana, e.g. concerning the search term "migration". The histogram on the timeline below the map will indicate how many items are registered in Europeana within a specific timeframe.

Mobile Access Channel
EuropeanaConnect will address the user demand to access information through mobile devices. The project will develop both a middleware layer, which will allow different mobile devices to access Europeana services created and a rich mobile client for real end user evaluation. The middleware will allow for access to Europeana from different mobile devices, channels and applications. It will be located between mobile devices on the user side and the Europeana semantic layer and database on the server side. It is intended to overcome the heterogeneity and limitations of mobile devices by selecting and adapting the content for queries to make them displayable on the devices. Therefore, queries from mobile devices are processed by the middleware, which uses the Europeana semantic layer to retrieve the desired data. The query results are collections of media which have to be adapted for mobile usage and a presentation of the results based on the metadata will be prepared. Finally, this presentation will be transferred to the mobile device. Advanced queries and interfaces require more powerful mobile devices, such as Smartphones. Therefore EuropeanaConnect will also implement a rich mobile client application which runs on Windows Mobile-based devices.

Value-Added Services for Europeana
EuropeanaConnect will integrate three value-added services with the Europeana portal that will significantly enhance the functionality and the usability of Europeana:

Multimedia Annotation Service
This service will allow all Europeana users to make their own contributions to Europeana in the form of tagging, comments, discussions, and linking. Annotations are an important means by which users can contribute knowledge and exchange information about content within a group of experts. Cultural institutions and museums that have already exposed their content on the Web are discovering that annotations can change the traditional role of visitors from passive consumers to actively contributing and collaborating users. Annotations allow users to exploit and add value to existing content, while enhancing accessibility through the user-provided searchable metadata and links to other media. Annotations further enable communication and collaboration with other users interested in the same content, which will support the establishment of social networks built around Europeana. The Annotation Service will allow the tagging, linking, and annotating of Europeana media resources of various formats (HTML, image, audio, and video).

GIS Service Suite
The GIS Service will allow users to query and display Europeana content based on spatial information, and to discover new relationships between content items, based on location. As only very few objects in Europeana already contain explicit geographical metadata, this service will also provide metadata enrichment for the spatio-temporal visual user interface described above. EuropeanaConnect will integrate a Geoparser and Gazetteer developed within the context of the DIGMAP and TELplus projects with the Europeana portal. By processing Europeana metadata with the Geoparser it will be possible to detect geographic terms and time periods. This will allow to improve the Europeana metadata. Pre-processed data will be stored in the semantic layer and will later be used to populate the spatio-temporal user interface. Finally, we can apply the same processing in real time for the users' queries. When the Geoparser finds a

relevant geographic or time reference (for example, "churches in the Scotland", "the battle of Lepanto", etc.) in a query, it can trigger specialised services that can take that in consideration.

eBooks-On Demand
Although Europeana provides access to some millions of digitised items from European cultural heritage institutions, these figures represent only a very small fraction of the analogue collections available in these institutions. In the case of books we know that the portion of already digitised material is around one percent of all books, in the case of archival material it will be even lower. This situation has been the starting point for the eBooks on Demand (EOD) Servicethat will be integrated into Europeana: The EOD Network (http://books2ebooks.eu/) is a growing consortium of 18 libraries from 10 European countries. They all provide the EOD service which allows users to order on-demand digitisation of public-domain books of the participating libraries and to receive the requested items electronically in the PDF format. EuropeanaConnect will adapt the EOD service allowing for an automated and generic transfer of the eBooks created within EOD to Europeana according to the specification of the Europeana Semantic Elements.

INFRASTRUCTURE COMPONENTS FOR EUROPEANA
EuropeanaConnect will implement and deploy key infrastructure components that will enable Europeana to manage the harvesting of thousands of digital heritage content sources across Europe, to promote metadata and content interoperability across content providers as well as interoperability with independent value-added services, and to provide persistent and uniform identification of digital resources:

- The Europeana OAI-PMH Management Infrastructure will be a prerequisite for enabling Europeana to meet the demands of the expected leap in data harvesting scale, managing the harvesting of huge volumes of content from thousands of content providers in Europe.
- The Europeana Metadata Registry will be a precondition for metadata interoperability in Europeana, as it will manage the range of metadata schemata and terms used by the various European cultural heritage institutions
- The Europeana Service Registry will enable Europeana to integrate practically any external service offering an HTTP interface into the portal. It will allow users to search and select value-added services and will provide the functionalities for the Europeana portal to invoke appropriate services depending on the context.
- The Europeana Resolution Discovery Service for Persistent Identifiers will allow for uniform and persistent identification of resources between Europeana and service and content providers. It will provide a crucial component for the portal's interoperability that supports the integration of a critical mass of European digital heritage content for the end user.

EUROPEANA LICENSING FRAMEWORK
Even though content of Europeana is likely to be in the public domain or with the rights cleared, it will be essential to provide the user with an accurate description of the rights attached to the content. Appropriate licensing policies for objects and metadata as well as suitable implementation methods are thus needed in order to create trust for all parties. EuropeanaConnect will create the Europeana Licensing Framework and will build the necessary tools for Europeana to establish and declare the rights status associated with Europeana content. In close collaboration with the Europeana Office and the Europeana v1.0 project, EuropeanaConnect will establish a core set of interoperable licenses that cover the rights information for objects in Europeana. The proposed Europeana Licensing Framework will be validated through a broad stakeholder consultation process.

AGGREGATION OF AUDIO CONTENT
Apart from building critical components for Europeana, EuropeanaConnect will also act as content aggregator: One of the shortcomings of the current Europeana prototype is its lack of a critical mass of non-textual content. EuropeanaConnect will add the music dimension to Europeana. As well as harvesting massive amounts of audio, the project will create, implement and evaluate access to right-free or wholly-owned music for all users of Europeana. EuropeanaConnect will integrate the framework created by the DISMARC eContentplus project (http://www.dismarc.org/) in Europeana and extend the technical, editorial and legal solutions of this platform. The project will provide the

necessary infrastructure for harvesting, analysing and storing of audio metadata as well as music audio provision to Europeana. There will be a continuous approach to and involvement of new audio archives for the duration of the entire project, using the established infrastructure and harvesting process. It is anticipated that over the period of the project approximately 150 audio archives will participate as content providers, leading to an estimated 200.000 music tracks to be provided to Europeana by month 30.

CONCLUSION

The launch of Europeana has yielded a high public interest. Within the next two years the current prototype has to be developed into a fully functional and scalable service that meets the high expectations of the user community. The project EuropeanaConnect which has been outlined in this paper will contribute to this goal on several levels. It will deliver key technical components which are essential for the realisation of Europeana as an interoperable and multilingual service. It will establish methodologies for analysis of user-interaction that will inform the implementation the Europeana portal is driven by user needs. Moreover it will deliver the Licensing Framework for Europeana and aggregate a critical mass of audio content.

Acknowledgement

EuropeanaConnect is co-funded by the European Commission in the eContentplus programme.

Abstract
This essay provides digital Cultural Heritage actors with perspectives about Online education—especially, the for-profit universities. Discussions cast these newest players on the educational scene into longer-range historical context—from the birth of the University, appearance of the printing press, and rise of the Mass Press to the ongoing Web revolution. The treatment provides a technical focus on their nature and operational concerns as uniquely digital offspring of the Web in contrast to similar operations in Cultural Heritage. Capitalistic and entrepreneurial viewpoints are also featured in examples from the American Public University System. As depicted, the APUS model calls for re-engineering library, bookstore, and press operations toward direct support of cost-effective, electronic classroom materials. The results suggest proactively adding similar course-specific engineering to Cultural Heritage's portfolio. The vision is a cooperative one of advanced Web-based applications, multi-institution combinations, and partnerships with the Online educators—all of which could contribute to the emergence of powerful Knowledge Communities on the frontier of learning.
Keywords: Online For-Profit Education, Electronic Course Materials, Knowledge Communities

Aldus [Manutius] is building up a library which has no other limits than the world itself (Erasmus, 1508)

Introduction

The presentation comes with cautions, but also a professed personal desire to explore future options. Online education and burgeoning for-profit universities are prime representatives of the new Information economy--but largely outside the orbit of digital Cultural Heritage. We will need some perspective. For-profits, for example, are uniquely of the Web. Internet applications may be similar, but you should understand that the viewpoints can differ fundamentally from Cultural Heritage. Realize too that these are capitalistic and highly entrepreneurial enterprises. Their cost/benefit analysis may be somewhat tempered for the sake of educational tradition, but assumptions and values that often subconsciously underlie Cultural Heritage may be challenged. Hence, these discussions will add a joint layer of historical background as necessary prelude to virtual community building.

A Fully Virtual Viewpoint

For-profit online universities are born of the Web. Such origins presuppose technological engagement - but are exclusively in the virtual. They do not include Cultural Heritage's primary allegiance and steward over physical documents and material culture artifacts. As indicated in the following brief potpourri, even overlapping Web concerns can thus differ markedly. Note: I approach the technical side from early training in Systems Analysis and running data processing installations before encountering work in Special Collections - as well as a couple of more recent, yet now out-of-date books on the Web. My operational bias is KIS (keep-it-simple) planning and letting the machines do the work whenever possible.

Deep Web v. Open Web: The Information Highway divides between free information on the Open Web and controlled, often pay-for access through the larger Deep Web. Given the financial implications and personal commitment to the Open Access Movement, the former is much preferred for the Onlines. Our programmatic focus is largely validating for "trusted" resources. Recognized Cultural Heritage sites make this easier and are much appreciated. Some in the Cultural Heritage side may be more interested in Deep Web operations. Outsourcing access is a desired goal, but a good deal of time is spent in pricing negotiations and contracts in what is a chaotic marketplace. In addition, we become involved in the details of Digital Rights Management (DRM) and Authentication routines - e.g., password controls, IP recognition, and proxy servers. One also expects heavy reliance on measurement and metrics, which come at unprecedented levels for the Cultural Heritage community.

Chiara Cirinnà and Maurizio Lunghi (edited by), *Cultural Heritage on line: Empowering users: an active role for user communities. Florence, 15th - 16th December 2009* ISBN 978-88-6453-184-7 (print) ISBN 978-88-6453-187-8 (online)

Digital Preservation: Although I have taught the subject and have ready recommendations, preservation does do not enter my current virtual consciousness. The absence of artifactual values, limitless exact reproductions, and inability to identify the original makes the term a misnomer in digital realms. In a key distinction, these institutions are not stewards. They generally do not own print or material culture treasures. Rather materials are merely leased or electronically borrowed for teaching and research purposes. Our management issues are ensuring access and basic readability on normal computer screens - scanning levels and DPI standards are not relevant for our type of studies.

Finding Aids v. Full Text: This is not really a question. Modern student clients could care less about Finding Aids and catalog entries. Students are simply spoiled by Google. They increasingly expect direct access to anything ever created. With the effective absence of the Series concept in favor of storage locations, even often touted EAD seems pleasantly naïve with a lot of work and of questionable value for my university's instructional purposes.

Metadata: Beyond basic identification information, human cataloguing or input of extensive metadata ala METS is an overblown consideration from a practitioner perspective. Anything that cannot be automatically ingested by the machine implies extra costs and is suspect.

Search Engines Considerations: All Web site producers should consider search engines as an audience and the major determinant in driving traffic. Despite the necessity of great care in design for navigation, search engines frequently tunnel around entry portals and directly to your contents. It thus helps educational institutions if Cultural Heritage materials are properly labelled for citation purposes. In lieu of advanced Bayesian mathematics and tuning relevancy engines - recommended design for native discovery normally features registration with DMOS, well constructed Title field, use of HTML hierarchical headings, and exchanging links with fellow repositories and institutions.

Section 508/Handicapped Student access: American universities taking Federal Student Aid are required to make accommodations for handicapped students. Thanks to the World Wide Web Consortium's (W3C) commitment to Universal Access Principles, compliance is typically rather simple for electronic textual materials. Cultural Heritage products, however, may need to be interpreted through optical character recognition (OCR) or word-processed transcripts (See JSTOR/Ithaka's new Decapod project for small institutions). Images call for parallel preparation - normally implemented through extensive use of the "Alt" descriptive tag. Media, especially the interactive variety, is more challenging, but we are exploring advances in voice recognition to simplify transcriptions, where those are not already provided.

Standards: Long-term solutions or standards from Cultural Heritage or traditional communities are of little interest. We look instead to the marketplace for answers. In Online Higher Education, SCORM compliance is expected for Learning Management System (LMS) operations. Some of us are also active in the ongoing development of Common Cartridge applications, which should draw your attention.

Transparency/Community Building: We strive to reduce the learning overhead and raise the comfort level on the Web for the students and faculty with the crucial subtext of building a lasting campus community. Overly jargoned or unnecessarily erudite sites and materials, which require extra hoops to retrieve, are not favourably received. Ideally, we would hope for continuity and seamlessness among Cultural Heritage sites. As will be reinforced later, technical goals extend with the Web's movement toward interactive customer services - ala the all-in-one, personalized "my portal" carrels. In our case, tailored virtual study arenas should be available including scheduling, classrooms, course materials, and social networking options. Amazon-like profiling and RSS feeds add to the future picture and could include Cultural Heritage materials - and communities.

Web & a Hyper-Incunabulum: Ultimately, it took Aldus Manutius in Venice - the inventor of portable books, italics, modern diacritics - and other printer/publishers the half-century or "Incunabulum" after Gutenberg to mature the book. The Web is engaged in a similar evolution. Permanence cannot enter our present calculations-planning logically extends no more than five years. At the moment, my staff is engaging mobile and other possible delivery platforms - like the Kindle

and Web PC. We also proactively engage communities wherever our students are congregating on the Web - e.g., Facebook, Twitter and on our island in Second Life. I even lecture on the rapid onset of Web 2.0 as a Third Wave of development in online education and stretching the boundaries beyond the LMS. Yet, those talks must also postulate the rapid onset of a Fourth Wave. Potentially transformational applications with voice recognition and, especially, touch screens are already in the queue. More importantly, a very different thinking and technically savvy Millennial generation is entering our colleges; it will likely have a Manutius or two.

Entrepreneurial Context/Example

The first publishers are often viewed as the first modern capitalists - they commoditized information and helped launch a Renaissance in the process. That role is being carried forward in the Web Age by online universities. Since the start of the 21st Century, online programs have been by far the fastest growing element in Higher Education. The sector has quickly grown to command some 10% of the market and is continuing to grow in excess of 12% a year. The prime example for this presentation, the American Public University System (APUS), is part of a handful of elite onlines that have earned regional accreditation - the *sine qua non* for American colleges. Since earning that status in 2006, APUS has been in hyper-growth. It now delivers Web-based courses to some 60,000 students in 150 countries at the bachelor and master's level. Moreover, the business has gone public with listing on the NASDAQ stock market (APEI symbol); hence, accountable to Wall Street and needing to speak differently than to archivists, curators, and librarians.

Accreditation standards aside, education in online universities differs in a number of ways from what the reader likely experienced. Ironically, Onlines contain throwbacks to the very origins of the university in 13th-Century Bologna. The dialectic again reigns, and every student must "speak" every week. Classes are small and not framed for the professor's schedule, but oriented to the convenience of the student client. Instead of two or three semesters a year, APUS launches 8 and 16 week sessions every month. Rather than having to go to a physical classroom and listen at set hours, the Web is the venue and asynchronous education the mode. Discussion groups can be picked up at any time of the day and any place with an Internet connection. Web-based instruction within LMS software brings unprecedented tracking and metrics on students and faculty alike. Our faculty are not physical campus assets, but scattered around the country and the globe. Moreover, the School's business model rigidly controls for pricing. APUS has not raised tuition in a decade and pioneered underwriting the entire costs for its undergraduate course materials.

APUS looks favorably on Cultural Heritage. Unlike many for-profits, the University has not abandoned the humanities or the arts; History remains among its largest programs. Rather than outsourcing or minimizing, the School is also dis-

The Idea of the University

John Henry Newman's famed *Idea of the University* marked an awkward transition. The Germanic-inspired 19th-Century New University Movement redefined Higher Education in a dramatic fashion away from his religious and classical ideals. In the United States, a general rise in literacy from mandatory grammar schools and the revolutionary appearance of a Mass Press set the stage for reforms that followed from Civil War-era Morrill Act. Centuries-long adherence to the liberal arts gave way to practical curriculum. Now familiar, albeit then new departmental structures appeared from agriculture and engineering to history and political science—along with budding professional presence for business and even librarianship.

As epitomized by the University of Illinois, monumental library structures appeared to define the heart of new campuses and as part-and-parcel of an overriding scientific research agenda. The library provided the primary market and storage place for outputs from new Ph.D. degrees with their publish-or-perish spate of scholarly journals and monographs. The Academic library shared the stage with newly invented public libraries, "the people's university" and rapid onset of museums and national libraries in the era—institutions characterized by a theoretical commitment research and development of taxonomies to control the rapid expansion in knowledge and print.

tinguished by significant and ongoing commitment to an academic library - a traditional seat for Cultural Heritage. Regionally or even national-distance accredited schools can have community service as an evaluation factor. In APUS's case, this factor includes outreach to the Cultural Heritage of its home base in West Virginia. Moreover, regionally accredited universities are inexorably linked to the established idea of the university - a concept that arose from the same romantic and nationalistic forces that effectively gave birth to the separate "cultural" sphere.

While linked to the historical lineage, APUS's specific version of the academic library could neither luxuriate in romantic assumptions, nor bask in a huge building with millions of volumes. It had to build anew within the context of the Web Revolution and competitive nature of a for-profit institution. APUS does not own, for example, but must follow a new mode by licensing e-books and electronic runs of scholarly journals. The new Online Library also defaulted any thought of holding all the world's knowledge to the Information Highway.
Our analysis revealed an inherent contradiction and opportunity from the past. With the exception of course reserves, classroom support was largely missing from the traditional scenario. American academic libraries would even eschew a role in the post-WWII democratization of education. They deferred to that era's textbook solution. Much to the confusion of generations of students, libraries would neither collect nor, if they had them, loan textbooks.
In a major paradigm shift, APUS's virtual facility and Collection Development were deliberately re-engineered for direct course support. Librarians were recruited as subject specialist to proactively create course-specific electronic bibliographies and support portals. The business goal was a 15 to 1 ROI (return on investment) through projected savings from electronic course materials. Entrepreneurship did not stop with the Library segment. The second leg of a tripartite strategy featured the Library absorbing and revamping bookstore operations along related lines. Rather than a pass-through, they became active bargaining agent in a drive to control rampant inflation in textbooks. The third deals similarly with contradictions in the university press. Instead of monographic lists, the APUS model refocuses a university press toward classroom support by commissioning select online textbooks and course packets. In sum, we present a new type of client for Cultural Heritage.

OPPORTUNITIES - ENGINEERING FOR EDUCATION

In addition to a fascination with the unfolding Web Revolution, let me admit to personal biases for Cultural Heritage efforts writ large. I come from a small island off Southern Louisiana. In my day, children were not allowed to speak French on the school yard. In the 1970s and 1980s, chefs, folk artists, musicians, government officials, scholars, and the people got together to instead celebrate their culture. Today, the term "Cajun" is internationally known, and I can find Louisiana-style restaurants around the world.
From a business orientation, however, Cultural Heritage products are a typical—albeit often technologically basic parts of the collection development spectrum. Providers are no different than the publishers and vendors that we deal with on a daily basis. Yet, their informational products are not like buying a car or computer—course materials help define the nature of our educational ventures. That factor along the uncertain nature of the Web and its "long tail" alter the simple buyer/seller relationship. Not unlike Erasmus working with Manutius in projects that helped launch personal authorship and humanistic scholarship, my office regularly engages in partnerships with publishers and vendors. The likes of Pearson, Wiley, Ebrary, and even JSTOR's Ithaka Sustainability Project seek our input and to collaborate with us in these challenging times.
Herein lies opportunity, but also the need for Cultural Heritage institutions to engineer a portion of their portfolio for classroom related services and Web product development. Standing still clearly will not work—especially as you digitize and publish results on the Web. Scholars, aficionados, and other institutions are already mining your resources. Yet, opportunities do exist for levels of engagement beyond particular collection, genre or provenance approaches driven from the Cultural Heritage institution. Although individual institutional efforts along these lines are laudable, especially those in the K-12 arena, they do not fully serve Higher Education purposes. A university focus turns on disciplinary requirements and teaching packets for specific courses from multiple resources and with multiple forms of Web delivery. My school would look for partnerships on our specific curriculum. In terms of applications on the immediate horizon, could we engage you in educational product development using budding Voice Recognition systems - or how could the new dynamics of touch screen applications add depth and different nuance to your materials?

The vision for Cultural Heritage pushes even higher. The Web's long tail beckons the evolution of complex Knowledge Communities as worldwide networked universities of the future. Individual and multi-institutional resources are marshalled around a subject, geographic area, or a people. To that base are added scholarly interpretative resources, but also schools of thought and opinion. Academics would be joined by curators, archivists and librarians with their own set of complimentary skills - as well as students and interested amateur. Discussion space is supplemented by wikis, simulations, and still unforeseen applications. Individually tailored and group spaces are enhanced by dynamic updating with new relevant monographs, journal articles, and discussion topics.

Conclusions

On one hand, the vision of Knowledge Communities is an unfolding challenge and an offer of cooperation. On the other hand, my conclusions do contain existential dangers. Unless carefully orchestrated, digitization and allied publication on the Web can make much of Cultural Heritage obsolete. This is the logical culmination of Manutius' drives and Trithemius' related lament *In Praise of Scribes.* With the transcendent appearance of the printing press 5,000 years of scribal cultural was effectively eliminated in short order. Intelligent educational packaging can help obviate that danger for Cultural Heritage in the Web Age.

That said, allow me to close on a very different note from my own time in Special Collections and Cultural Heritage before being recruited to the world of online education. In the late 1990s, I was visiting Jac Treanor, archivist of the Archdiocese of Chicago, along with Monsignor Charles Burns - then head of the Vatican Secret Archives. Discussions turned to labor negotiations that I was having when Burns piped in with his Scottish lilt. Just before leaving Rome, he had been similarly engaged - albeit through an older work-related document. St. Peter's Cathedral was under construction, but the Cardinals were in conclave and the project running out of money. The foreman passed a note in. "For the sake of St. Peter" (purportedly the origins of the aphorism - "for Pete's sake), if funds are not made available, I will need to shut down this job. Signed: *Michelangelo*. Please keep in mind that there is no way that my beloved Web and digitization can duplicate the resulting "Michelangelo Effect" - with its visceral, associational and tactile pleasures.

Abstract
Following EC´s i2010 strategy European libraries are currently systematically digitising and making available their cultural heritage. Very often, however, researchers, readers and users demand access to historical books „here and now". This is exactly where the eBooks-on-Demand (EOD) network starts from, providing a trans-European digital document delivery service for end-users from all over the world. Currently the EOD network comprises over 20 libraries from 10 European countries. Since 2007 several thousand PDF eBooks have been produced, delivered to users from over 30 countries and made available to the public through participating digital libraries. User reactions are very positive and more and more libraries are interested in the service. This paper is about the following topics: 1) the user-driven approach of content selection for digitisation, 2) financing such a service, 3) results from a user survey, 4) future perspectives within supplying web 2.0 platforms.
Keywords: Digitisation, eBooks on Demand, user selection

Introduction

In line with EC`s i2010 strategy European libraries and other cultural institutions are currently systematically digitising and making available their cultural heritage to a wider pubic. Due to the enormous amount of material, it will take some decades until all books, journals and other library material will be available in digital form. Ideally, all the works in a collection or holding should be considered for digitisation. In practice, however, this is rarely feasible and choices must be made. According to the MINERVA Good Practices Handbook (Clissmann 2004) the following criteria may be suggested for setting preferences: 1) Access to material which would otherwise be unavailable or of limited availability, 2) wider and easier access to very popular material, 3) condition of the originals, 4) preservation of delicate originals by making digital versions available as an alternative, 5) project theme, 6) copyright and IPR, 7) availability of existing digital versions, 8) cost of digitisation and 9) appropriateness of the source material for online viewing.
In many cases the chosen approach is determined by the source of funding, preferably in combination with the institution's digitisation policy. This way of processing digitisation only very rarely acknowledges interests of users apart from point 2) above which highlights popularity as a criterion of selection. This raises the following questions: 1) how can individual users´ needs been considered in digitisation?, 2) what happens to and who will take care of those materials not covered by all those projects?, 3) more precisely, who will digitise books of minor languages or those from special or smaller collections or institutions?
All those considerations formed the starting point for a service called "eBooks on demand": an electronic document delivery service instigated and co-funded by the user. What was needed was a model for ongoing, permanent digitisation, a coordinated European initiative where selection of material necessarily works bottom-up rather than top-down and which is instigated by the individual reader.
Experience has shown that such a project needs co-financing by the reader requesting a certain book to be digitised. By saying so I only refer to the co-financing of scanning itself. The important but often neglected issue of subsequent costs such as storage, access, long term preservation and migration costs will not be touched here.

EOD network and service

Of course, such an additional library service can't be maintained within the context of usual (mass) digitisation only, but needs extra resources for order management, customer communication, payment procedures and so on. Therefore, a structure allowing for efficient processing of orders was designed. At the end a central-decentral service network was implemented, where some processes such as OCR or online payment are hosted centrally. From October 2006 to June 2008 a pilot project and evaluation was carried out within the framework of "Digitisation on Demand", co-funded by the

Chiara Cirinnà and Maurizio Lunghi (edited by), *Cultural Heritage on line: Empowering users: an active role for user communities. Florence, 15th - 16th December 2009* ISBN 978-88-6453-184-7 (print) ISBN 978-88-6453-187-8 (online)

eTEN programme, consisting of 13 libraries from 8 European countries. By July 2008, a self-sustained network was up and running with these 13 libraries as founding members.
In the following I will briefly describe how EOD works. The starting point is the online catalogue of any participating library. There, the EOD button is placed with all items available for digitisation. At the moment, these are books which fulfil at least one of the following criteria: 1) items not yet digitised, and 2) public domain books. Any user interested in a certain book tagged with the EOD button simply needs to click on this button. The library then receives the order, scans the requested book and transfers the replica via FTP to the central server located at the network coordinator Innsbruck University. Each library manages and processes the orders in a central database accessible via a normal web browser. Here, also the digital object is created. After the completion of the payment process, the user downloads the PDF from his personal tracking page. After some time the library eventually integrates the images into its digital library or repository and thus makes them available to the general public.
From a library's point-of-view the EOD service is a mixture of central and peripheral services sharing two core components: The order data manager and the digital object generator. For its own orders only, each library has access to the Order Data Manager, a central database with web-interface. According to its needs and objectives any participant is able to adapt and customize the website texts, automatically generated email texts, etc. Core part of the Order Data Manager is the Digital Object Generator, a central service for generating eBooks. It allows OCR recognition (antiqua and gothic), automated cover generation, PDF & RTF delivery, creation of Abbyy XML and the generation of the streaming link for downloading. Online payment in the form of credit cards also forms part of the central services – so the individual library doesn't need to grapple with business like this.
Currently, more than 20 libraries from 10 European countries are offering the service, ranging from Portugal in the far west to Estonia in the East. The University of Innsbruck library is both, the coordinator and central service provider of this network at the same time. Concerning the types of libraries involved - there are 6 national libraries, several university and state libraries as well as research and academy of sciences` libraries.

Austria	**Czech Republic**
University and Regional Library of Tyrol (co-ordinator)	Moravian Library in Brno Research Library in Olomouc
University Libraries of Graz and Vienna, Library of the Medical University of Vienna	Library of the Academy of Sciences in Prague
Vienna City Library	National Technical Library
Germany	**Hungary**
University Libraries of Regensburg, Greifswald and Humboldt-Universität zu Berlin	National Széchényi Library of Hungary
Bavarian State Library Saxon State and University Library (Dresden)	Library of the Hungarian Academy of Sciences
Denmark	**Portugal**
The Royal Library	National Library
Estonia	**Slovakia**
National Library University Library of Tartu	University Library in Bratislava
France	**Slovenia**
Medical and Dental Academic Library of Paris	National and University Library

Table 1: Overview over EOD libraries

Pricing and user experience

Up to now some 3200 books have been digitised and delivered to customers - books that wouldn't have been otherwise digitised. This makes up some 840.000 scanned pages for nearly 2000 users worldwide. The top 3 libraries

receive about one request per working day and deliver 250 to 350 books per year. The average delivery time is one week.
The average price of an order is about 50 EUR. The price of an eBook is set by the respective library and is calculated from a base price plus the number of the pages. There is still a quite heterogeneous pricing system as at the moment each library sets up its own prices. In average, digitisation via EOD includes a basic fee of 10 Euro plus 0.15 to 0.30 Euros per page. In view of this price range, a certain degree of harmonization certainly needs to be agreed upon in the near future.
Experience has shown that the revenue gained from digitisation merely co-funds digitisation – to be more precise: the actual personnel costs for scanning including additional costs for the service as such. Quite clearly then a customer-oriented service leads to higher costs than mass digitisation, which lays huge emphasis on efficiency and economisation. Given, though, that the master files remain with the library and is later made freely accessible to the public, the user by no means can be made to pay the actual total cost of digitisation. The user must be seen as the instigator and co-funder of digitisation; in turn the library profits by building up its digital library in the long run.
In 2008 we carried out a user survey among EOD customers to receive feedback to the service. We also wanted to find out what users thought about our pricing policy. The findings were that 30 % of our customers thought the price was high or very high, but still the overall price-value relation was found acceptable by the majority.

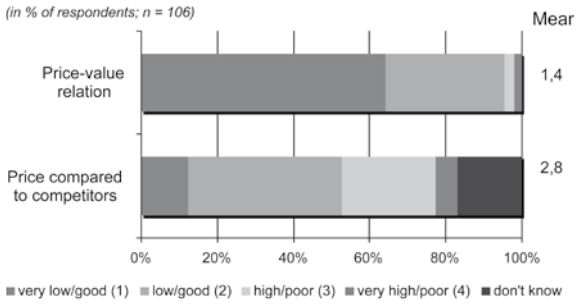


Figure 1: Users' evaluation of price-value relation

Its noteworthy that the overwhelming majority of users are either researchers or people who require eBooks for "professional or scientific use" (over 60 percent). Second place (16 percent) is book collectors and people who could be said to count among special interest groups such as amateur historians, collectors, or ethnographers for example.

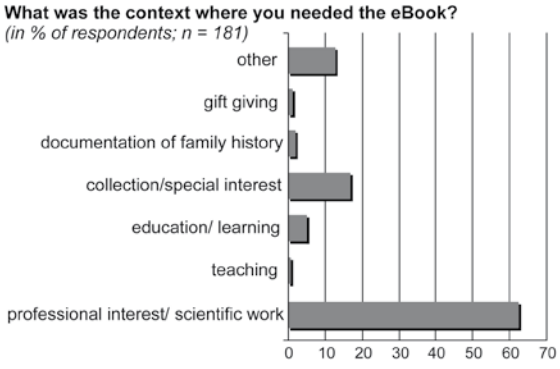


Figure 2: Users' domain of interest

Asked why they chose the EOD service, almost half of the interviewees answered that without EOD the book would have been "impossible or difficult to access". This shows that EOD achieved one of its main goals of being an additional alternative for accessing books.

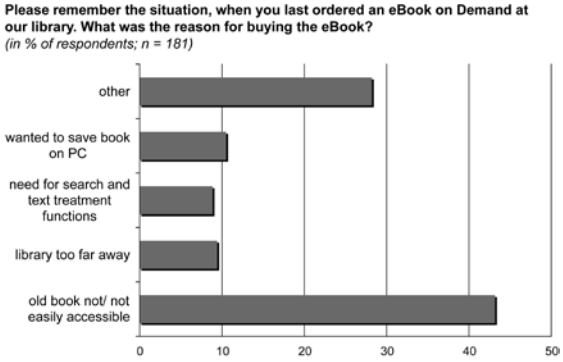


Figure 3: Reasons for ordering

In the EOD customer survey, 60% of our customers told us that they usually print out selected pages or even the whole book on paper. Thus, there was an obvious request for "re-materialization" of digital material. In response, we have lately been offering "reprints on demand", supplementing the digital file. This service is also carried out in a centralised way. Each library only needs to take care of scanning images and some more metadata. Everything else connected with the service such as image enhancement, the creation of pre-press PDF and related files is carried out by the central coordinator at the University of Innsbruck library.

Future perspectives

Within the EC Culture programme a new project was initiated in May 2009 with duration of 4 years. The project will focus on larger scale involvement of three main target groups: participating libraries, requesting users and the general public. 20 libraries from 10 European countries take part in this project which mainly concentrates on the following 3 objectives: 1) to enlarge the EOD network by additional European libraries, especially those from countries not yet represented 2) to take EOD as a best practice model for any other European-wide network of this kind; furthermore to train stakeholders (libraries, museums, or other cultural operators) to run such a multinational cultural service based on state-of-the-art information technologies and 3) to support intercultural dialogue among readers and users of historical books with the help of web 2.0 technology.

Within this latter goal web 2.0 based social platforms such as Wikipedia, LibraryThing or Goodreads will be supplied with information on selected historical books. Readers all over the world interested in a specific book thus will be able to easily interact with each other, exchange information and share reading and research experiences – independently from where they are and which background they may have.

During the last years the internet has become more and more of a "social network". Cultural organisations need not only "offer" their cultural heritage in the internet, but they need to pro-actively "supply" web 2.0 platforms with their content, since these are the sites where the majority of users are actually present. Not only do users create and edit information, but upload texts, images, and videos, they "meet" other users, they "share" content, they "cooperate", "rate" and generally "take part" in a digitally mediated community. Whereas many sites, such as library catalogues or repositories, are only visited sporadically in the event of retrieving some clearly defined data, people have their personal accounts at web 2.0 applications and use them regularly – comparable to the way they regularly read a newspaper, meet friends at a cafe or visit a museum. First, EOD will therefore explore the technical feasibility to automatically generate Wikipedia source-code for selected books which have been ordered and digitised via the EOD service to then

later on add metadata to articles about the respective authors. Other social platforms will be tackled in a second step.

CONCLUSIONS

Experiences made so far indicate that "on demand" services are desirable ways to make books accessible either in digital or in re-printed form. EOD has proven to be an important additional library service, making holdings all over Europe more accessible to customers and providing researchers with material they need "on demand", here and now. In fact more "on demand" services could be imagined, such as digitisation on demand for the blind and visually impaired and the creation of „real" eBooks with a corrected full text approximating 99% accuracy. This would pave the way to transfer eBooks onto mobile-devices, making the written word available every time, everywhere.

ACKNOWLEDGEMENT

This work programme has been funded with support from the European Commission. This publication reflects the views only of the author, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

REFERENCES

[1] Clissmann, Ciaran (2004): Good practices handbook. Version 1.3. Minerva Project. Minerva Working Group 6 Identification of good practices and competence centres. Online: http://www.minervaeurope.org/structure/workinggroups/goodpract/document/goodpractices1_3.pdf [Access: 25/10/2009]

[2] Mezö, Zoltán; Svoljsak, Sonja; Gstrein, Silvia (2007): EOD - European network of libraries for eBooks on demand. In: Research and advanced technology for digital libraries, p. 570–572.

[3] Mühlberger, Günter / Gstrein, Silvia (2009): eBooks on Demand (EOD): a European digitisation service. IFLA Journal (35/1), 35–43.

[4] Online: http://archive.ifla.org/V/iflaj/IFLA-Journal-1-2009.pdf [Access: 25/10/2009]

[5] Svoljsak, Sonja; Gstrein, Silvia (2007): EOD - eBooks on demand. In: Users and use of DL & economics of digital libraries, p. 229.

Abstract
By utilising web-based collaboration tools, institutions can engage users in the building of historical printed text resources created by mass digitisation projects. The paper presents the drivers for developing such tools and identifies the benefits that can be derived by both the user community and cultural heritage institutions. The perceived risks, such as errors introduced by the users or whether users will engage with resources in this way, will be set out. The paper will present the lessons that can be learnt from existing activities, such as the National Library of Australia's newspaper website, which supports collaborative correction of Optical Character Recognition (OCR) output.
The user collaboration tools being created by the IMPACT Project (Improving Access to Text, http://www.impact-project.eu), a large-scale integrating project funded by the European Commission as part of the Seventh Framework Programme (FP7), will be detailed. A primary aim of IMPACT is to develop tools that help improve OCR results for historical printed texts, specifically those works published before the industrial production of books in the middle of the 19th century.
While technological improvements to image processing and OCR engine technology are key to improving access to historic text, engaging the user community also has an important role to play. Utilising the user community can aid in achieving the levels of accuracy currently found in born digital materials. Improving OCR results to this level is key to producing resources that support better resource discovery and enable greater performance when applying text mining and accessibility tools to the extracted text. The IMPACT project will specifically develop a tool that supports collaborative correction and validation of OCR results and to allow user involvement in building historical dictionaries that can be used to validate word recognition. The technologies use the characteristics of human perception as a basis for error detection.
Keywords: Digitisation, OCR, User Collaboration, IMPACT

Introduction

The Challenge

The digitisation of historical text resources and use of sophisticated software tools to translate the images of text into machine-readable text has transformed the way researchers engage with these types of resources. CENL (Conference for European National Librarians) surveyed its members in 2008, revealing an expectation of a 350% increase in the digitisation of historical books and newspapers between 2006 and 2012, which would make them the most popular type of material being digitised[1]. The benefits of OCR (Optical Character Recognition) in the digitisation workflow were recognised but the experience at the British Library in their project to digitise 19th Century newspapers indicated that there were issues in the quality of the OCR text with, on average, over 20% of the text on a page not being correctly recognised [2]. Many factors influence poor performance, such as quality of the original material, storage practices, and the fonts and languages used. Current OCR is tuned to processing modern printed text and so there is a need to improve the performance of OCR tools when dealing with historical texts.
Achieving 100% percent or even the 99.9% accuracy that is usually specified through completely automated solutions will therefore be difficult to achieve for historical text. Indeed, when accuracies of this level are required then the preferred solution is to get human operators to re-key the data. While this produces good results, it is not scaleable and when institutions are digitising millions of pages the costs are prohibitively high. The cognitive ability of the humans make them ideally suited to task of recognising text that computers cannot. So there is a need to harness that power in a way that can scale to support projects that are digitising millions of pages of text. The ideal solution is to partner the strengths of the computer and humans to achieve our goals for accuracy.

Chiara Cirinnà and Maurizio Lunghi (edited by), *Cultural Heritage on line: Empowering users: an active role for user communities. Florence, 15th - 16th December 2009* ISBN 978-88-6453-184-7 (print) ISBN 978-88-6453-187-8 (online)

THE POWER TO BE HARNESSED

The advent of technologies which carry the moniker Web 2.0 technologies, the interactive web, has meant users are not just presented with information but play a full part in the creation, enhancement and semantic mark-up of information. Amazon, Wikipedia, Youtube and Facebook all provide evidence that the web community has fully embraced this paradigm, indeed Flickr recently announced that the 4 billionth photo had been uploaded [3].
The ability to harness the millions of users who interact with web-based cultural heritage resources to fill that gaps that OCR software leaves behind is attractive, but there remain questions as to whether users will engage to the same extent in collaborative correction.

The concept of user collaboration in the clean-up of OCR text is not a new one. Distributed Proofreaders have used a volunteer force of over 4,000 people to correct the OCR text of over 16,000 titles in a nine-year period [4]. While this approach demonstrates that users are willing to engage in such activity for the benefit of the community, it is not a mainstream activity routinely deployed in the digitisation workflow of cultural heritage institutions. There is a need to significantly increase user throughput to match the levels of digitisation that is currently under way.
Two recent initiatives provide further demonstrable proof that user collaboration is a key tool to resolving the accuracy gap over purely computational approaches. Indeed a fusion of the computational approach of modern computer software allied with the cognitive power of the human brain can help us achieve the results we seek.

CASE STUDIES

reCAPTCHA [5], which has a tag line of "Stop spam, Read books", is the result of a project undertaken by the School of Computer Science based at Carnegie Mellon University. CAPTCHA stands for Completely Automated Turing Test to Tell Computers and Humans Apart, and are text forms used by websites to prevent automated programs from accessing websites for malicious purposes. reCAPTCHA uses computer-unrecognised OCR results for these text forms, and thereby uses human interaction to improve the initial OCR results for historical documents. There are no published statistics on the number of words that reCAPTCHA has corrected but over 200 million CAPTCHAs are solved every day [5] and almost one billion reCAPTCHAs were solved in 2007/2008[6].
In July 2008 the National Library of Australia released a public beta of the Australian Newspaper service [7] which supported public collaborative OCR text correction. It was the first service of its kind. It was a low-key launch with a desire to get feedback from early users to see what they required from such a service and how they would engage with the service. One part of this approach was to leave moderation of user changes to the community.
In the first 6 months of the project the usage of the service was as follows:

- 2,994 registered users
- 2.2 million lines of text corrected
- 104,000 articles corrected

The verdict of both the National Library of Australia and the site's users on the first 6 months of the service is overwhelmingly positive. The issue of users making malicious changes did not surface and no vandalism of text was detected. Indeed, users validated the community moderation approach by explicitly stating that users should moderate each other, rather than the Library moderating, and have the ability to report and correct issues.
These two examples demonstrate that user communities are a power to be harnessed in improving the quality of OCR text.
The IMPACT approach to Collaborative Correction
IMPACT (Improving Access to Text, http://www.impact-project.eu) is a large-scale integrating project funded by the European Commission as part of the Seventh Framework Programme (FP7). The project commenced in January 2008 with the following objectives:

- Develop OCR software and technologies which exceed the accurateness of current software significantly.
- Provide a software system which will allow the realisation of new concepts of collaborative correction (in order to lower the costs for full featured full-text) by taking up and integrating Web2.0 concepts
- Develop language tools and lexica in order to provide access to historical texts independently of historical variants of a given language.

- Support adopters of these tools so that more European historical lexica can be built.
- Develop a number of smaller modules such as image enhancement and segmentation toolkits, functional parsers, etc. in order to support the automated text recognition and/or access to historical text.

It was recognised that while the project could seek to advance the state-of-the-art for language and OCR tools, there was a need to provide advanced solutions in engaging users to improve the word accuracy of digitised historical texts. The context for this is the i2010 vision of a European Digital Library: an ambitious plan for large scale digitisation projects that will transform Europe's printed heritage into digitally available resources.
Building on the experience of such initiatives as reCAPTCHA and the Australian National Library's Newspaper project, IMPACT will develop an alternative approach to user collaboration in the clean-up of OCR text which will increase the power of these types of tools.
The tools will allow for involvement of the general public in validation and correction of OCR results. These tools will be based on the SmartKey idea described below. This technology uses the characteristics of human perception as a basis for error detection. The result is a data acquisition procedure that is very efficient and virtually error-free.
The OCR engine recognition process concludes with a rating score for each character. These scores are further refined by the spell checker, which can either increase or decrease the probable success of individual characters. Based on such probabilities, all the characters are classified as Sure (characters with a high enough probability), Medium or Unsure (characters with a low enough probability). While Sure characters need no further verification and can be accepted automatically, and Unsure characters are sent for manual data entry, the Medium characters are sent for fast verification via "carpets".
In a "carpets session", all the Medium characters from different sources (possibly different pages, chapters or even books) are sorted in alphabetical order. For example, all the characters that were recognised as an 'A', but with a low score rate, are grouped together in a single "carpet". It has been shown to be very easy to identify the few errors and thus automatically approve the other characters as valid 'A's. Hence, instead of keying in 100 characters, it is sufficient to point the mouse at few errors and the others will be automatically deduced. As a result, the validation process becomes very fast and effective. Figure 1 provides an example of a carpet session were the user is asked to identify all items which are not the letter e.

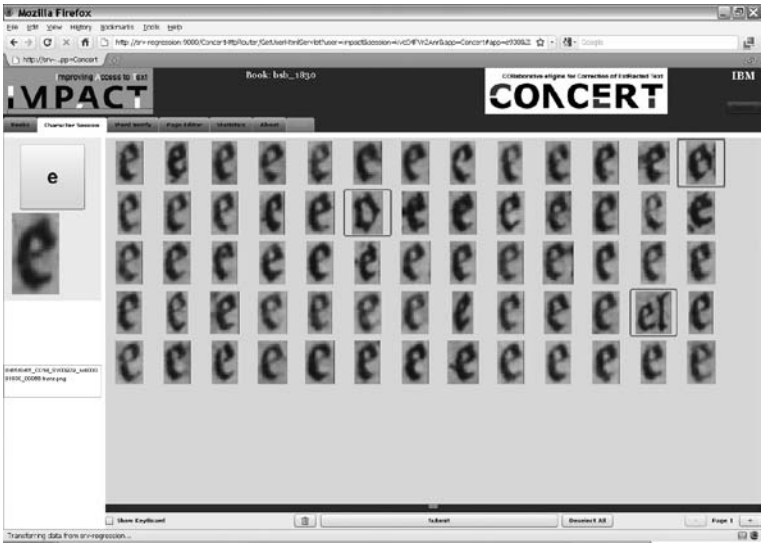


Figure 1 A View of a carpet session

At the end of this process, some words may remain unrecognised. Indeed, if image quality is very poor, it maybe impossible to recognise a given character without the context of the entire word or sentence. For these infrequent cases, word-based data entry will be introduced and context information made available as necessary. In this way the user can also add words to the dictionary supporting the OCR process by identifying words which are not currently included. This dictionary is part of an Adaptive OCR Engine and user collaboration will therefore not only correct words but train and enhance the engine's vocabulary and language analysis features. The system will dynamically decide how much manual intervention is needed to achieve a certain level of accuracy.
Quality monitoring will be enabled by feeding known errors into user "carpets" and seeing whether the user detects those errors. If the user doesn't, their results will be weighted as less accurate than those of someone who does. In extreme cases, the user may be defined as a malefactor and his/her contribution discarded altogether. Quality monitoring will be done online to facilitate the adaptive utilisation of the results.
To summarise, the IMPACT approach to collaborative correction involves:

- creation of a data validation/correction application that is simple and intuitive enough to be attractive to untrained users and yet effective enough to ensure high productivity
- developing a powerful control system capable of analysing and segmenting books and other documents into individual small jobs, ensuring successful job completion, and reassembling the final result
- creating a web-based application suitable for mass volunteer participation.

SUMMARY

Large-scale digitisation of historical printed text resources is a reality and is transforming the research landscape for this type of material. With the research benefits there are weaknesses, a key one being the quality of OCR text.

Advances have been made and will continue to be made in advancing the state-of-the-art in the automated translation of printed text into machine-readable text, and IMPACT is active in this area.
Significant benefit can be derived by using humans to bridge the gap between what can be done in an automated way and desired accuracy levels. Examples of this have been deployed with great success and will significantly improve digitised resources. IMPACT is introducing a new paradigm in this area, where not only is the text corrected but the user input is used to improve future OCR and language processing in a seamless manner.
Material that is born digital can have extremely high levels of accuracy, user collaboration will allow us to approach those levels of accuracy for historical texts in a cost effective manner.

REFERENCES

[1] Hans Petschar, et al. "EDL Report on Digitisation in European National Libraries 2006-2012" February 2008. Available at < http://www.cenl.org/docs/report_digitisation_nls.pdf >
[2] Simon Tanner, Trevor Muñoz, Pich Hemy Ros. "Measuring Mass Text Digitization Quality and Usefulness" D-Lib Magazine. July/August 2009, vol. 15 no 7/8 < http://www.dlib.org/dlib/july09/munoz/07munoz.html >
[3] < http://blog.flickr.net/2009/10/12/4000000000/ >
[4] < http://www.pgdp.net/c/ >
[5] < http://recaptcha.net/learnmore.html >
[6] < http://markmail.org/message/orklsq2e4csqr2s4 >
[7] Holley, Rose (2009) "Many Hands Make Light Work: Public Collaborative Text Correction in Australian Historic Newspapers." ISBN 978-0-642-27694-0. Available at <http://www.nla.gov.au/ndp/project_details/documents/ANDP_ManyHands.pdf>

Abstract
EHPS, [http://primary-sources.eui.eu], is a portal serving a community of Ph.D. researchers and post-doc fellows and professors at the Department of History and Civilisation its users with an easily searchable index of multi-lingual collections of scholarly websites that offer online access to digitised primary sources, invented archives and born digital sources relating to the history of Europe, either as a whole or for individual countries. EHPS offers web 2.0. features to remain connected to the portal and be informed about new entries.
Keywords: Portal, Digital History, Primary Source Collections, E-resources, Digital libraries, Meta-sources, Born digital Sources, History of Europe, European University Institute.

European History Primary Sources EHPS Home page

Introduction

This is the time for a new craft and a new methodology for a digital historian, a more common and day to day «pratique de l'histoire» as Marc Bloch would have said also using socially connected new web 2.0. features [1]. Historians and the communities to which they refer, need to organize better in their computers and through their browsers, the access to digital repositories and archives, manage data's through appropriate software's, archive in durable and secure ways their digital documents and artifacts and diffuse their writings in OA repositories.
Important digital archives and invented archives, research tools, digital libraries and other e-resources are available online. One may say that looking at these is now part of the process of selecting the necessary materials to write an historical essay. But new methodological issues are arising, not only for who's creating these new web-sites, invented archives and primary source libraries from physical materials, but also for who's using them for research and the writing of history both in a traditional way or again for the web.
The library supporting the research activities of historians at the EUI is not old and has no old collections of books and primary sources. It never even had a card catalogue but always had a computerized database of bibliographical records [2]. Such a small academic library couldn't afford to deal with a massive selective project of digitization for primary sources collections even for the early XX century. In the 1980s the library was concentrated on the purchase

Chiara Cirinnà and Maurizio Lunghi (edited by), *Cultural Heritage on line: Empowering users: an active role for user communities. Florence, 15th - 16th December 2009* ISBN 978-88-6453-184-7 (print) ISBN 978-88-6453-187-8 (online)

of primary sources in microforms and, later, when the web developed itself, had to face the financial situation academic libraries are now trying to deal with coping with subscribing and purchasing digital materials, mainly in English for his multi-lingual and multi-national public.
So, accessing digital primary sources at the EUI is accomplished locally, retrieving the microform documents, going on missions to libraries and archives abroad or using the internet and a browser based procedure. EHPS has been created in this context to avoid -when possible- the use of costly and time-consuming missions abroad.

A Digital Turn in History

Today, after the first important primary sources digitization projects in the 1990s in North America and Europe, we face new scenarios in history because the digital turn is now offfereing an abundance of digital primary sources worldwide. This is happening even if the American historian, one of the fathers of digital history, Roy Rosenzweig, challenged this idea of abundance: the immediate volatility and disappearance of new digital media formats and web-sites were challenging that assumption he wrote [3].
Kirsten Sword, professor of American women's history at Indiana University, suggest to use these new digital archives which are de facto offering an enormous potential for new historical inquiry: "digital resources are expanding and redefining the archival base for most fields and thereby redefining the fields themselves, she said even if "this is driven more by libraries and the tech industry than by historians" [4] and, during the same round-table about the "promises of digital history", Steven Mintz, past president of H-Net, wrote about what I would personally suggest to call a real digital turn in historical practices, that modified the way to make history: "it has greatly expanded the range of sources —primary and secondary— that I use…". [5]
During an international conference on Contemporary History in the Digital Age organized in Luxembourg in October 2009 [6], Marin Dacos, director of the Cléo, Centre pour l'édition électronique ouverte in the CNRS in France [7] (he founded in 1999 the project Revues.org) [8], insisted in his keynote speech that if a History 2.0 has to be defined, it is because we are building new cyber-networks for accessing our sources. He underlined the fact that our goal was to bring scholarly literature and the primary sources to everybody's own computer. This has to be organized in a open society model through web based 2.0 technologies and bypassing as much as possible, commercially owned technologies, commercial databases and private actors.

Small scale Primary Source Collections in a fragmented internet

It is a fact that from the end of the 20th century, scholarly digital contents is each day more in the hand of few commercial e-actors which used the intellectual production of scholars to create new revenues from online access to their own scholarly contents. Essays written by academics were sold back to their libraries and their universities.
A Web 2.0. debate "avant-la-lettre" was organized online in 2001 by Noga Arikha and Gloria Origgi for the Centre Georges Pompidou in Paris. Eminent scholars dealing with the online cultural heritage participated [9] . The historian Theodore Zeldin wrote a paper on "The Future of Internet". During the online discussion about the paper, I asked the author of a history of an intimate history of humanity [10], about the growing internal fragmentation of the web and about the parallel existence of different levels of access to web contents. I summarized this closure of the access to web contents in French as "internet et les internet(s)" asking him what could still be done for supporting the free diffusion and access to scholarly materials –primary sources and secondary literature- in the humanities [11].
One of the possible response to such a commercial challenge, is that huge digital projects of primary sources digitization in history are also coupled with small scale projects. These new projects are often using new 2.0 web technologies to attract the user's capacity to add contents or refine existing description of contents. These projects are extremely useful to digital historian activities organizing better the interoperability and integration with their research needs. Digitizing few sources or a single collection of sources is also done today within smaller cultural institutions, to concentrate online on more specialized contents with high scientific standards and high technological added value. This is done allowing good retrieval procedures within the digitized documentation also for remote uses.
This new trend to support digitization of less pharaonic projects, with more precise and delimited contents and freely accessible, is a new recommended policy to support digital history activities. Edward L.Ayers, an American pioneer digital historian, realized between 1991 and 1993 a SGML project called «The Valley of the Shadow, two communities

in the American Civil War» which became in 1996 a web project. After more than ten years of technical developments and the addition of an enormous amount of digitized primary sources about the Civil War in the USA, Ayers is now suggesting us to change our agenda and not to launch such large scale projects: «we've tended to build big things in the hopes of capturing as many uses as possible. But maybe now we need to build lighter, smaller things. We might build simpler ways to use our vast collections." [12]

The stages of the historian's workshop

Historians confronted with the digital age and digital history, even for those not thinking at all of producing digital artifacts within a Ph.D. research, needs to adopt a new digital craft made of new methodologies and new critical paradigms [13]. This new "workshop" is made of different practical and methodological working stages interacting with the internet and the web. If we could divide the process of doing digital history, four different steps would describe a kind of working process with which an historian would be confronted starting from the "production" of the digital documentation up to an individual "consumption" of digital primary sources. This process of research summarizes the history digital turn and defines the practice of digital history. All the four different activities below, are each singularly and together too, defining the field of digital history as a way to produce history using the new media and digital primary sources now available on the web.

- Production of e-sources;
- Information about e-sources;
- Selection and evaluation of e-sources and meta-sources
- Use of e-sources to write history in a traditional way or again for the web.

EHPS: monitoring primary sources websites and collections

But how to monitor, list, collect and organize the research of these scientific primary sources dealing with the history of Europe ? The European History Primary Sources portal tries to answer to this question dealing with the specific needs of a European community of historians based in Florence, Italy, at the European University Institute [14]. EHPS integrates itself within this new "digital turn" when not only digital historians will have to look for their documentation in the web because of the many added values the internet will offer them outside the traditional ways of dealing with primary evidences.

EHPS inaugurated officially in June 2009. While not claiming to be complete, it contains the major national digital libraries and many smaller series of e-sources and smaller digitization projects in Europe in all national and sub-national languages. It thus reflects in a way, the current state of digitization of historical source materials in Europe, as well as those digitized outside Europe pertaining to its history. The portal is aiming at tracing a map of all the different digital libraries and primary sources collections and databases available for the history of Europe from medieval times to nowadays in all European languages. EHPS precisely enters the second stage of the digital historian's journey, because its main goal is about informing of the existence of digitized primary sources and trying to offer its users a way to collect them and access them inside their browser and for further evaluation and uses, steps 3 and 4 of that same journey. EHPS was born as an autonomous part of the galaxy of web-sites belonging to the World Wide Web Virtual Library History Central Catalogue moved in 2004 from the University of Kansas to the European University Institute in Florence [15].

It responds, within its selected contents, to the question of how we could link better the process of searching for primary sources in the digital age thanks to History 2.0. services. In this way, EHPS is a complement of the new ways offered by new semantic OPACs for searching inside library holdings and combining internal search with external potential contents. EHPS follows also the economic necessity of adopting low technological profile solutions because few financial resources were available and no specialized ICT staff was available. EHPS is made of a very light open source CMS cyber-infrastructure, easy to build and develop, easy to understand and use. EHPS is a small scale tool, low technological profile database using Dries Buytaert's Drupal open CMS [16], and the Zen theme system maintained by John Albin Wilkins [17].

EHPS has been conceived as a web 2.0 tool aiming to win the participation of its specialized public in order to complete the single web-sites descriptions and abstracts it offers with personal experiences within these archives. It is hoped that EHPS would be considered not only as a redirecting device to other web-sites, but also as a research tool that

needs critical judgments from its users for the benefit of other users. This would be also about collaborative tagging and creating more articulated folksonomies completing EHPS keywords.
EHPS portal is a tool for indirectly «publishing» primary sources in the browser from a request of a fully digitized single newspaper like L'Unità in Italy or the Journal de Genève in Switzerland, to broader research topic like the Cold War or the European Integration, to a request for specific types of primary sources like posters or postcards for the history of the UK during the Victorian period, etc.. For postgraduate programs, the need is to discover original and unexploited primary sources or to consult and precise the use of original sources.

SEARCHING PRIMARY SOURCES IN EHPS

Within EHPS, primary sources are to be retrieved and accessed in two steps. The first step is made of a retrieval of the portal's contents, the second, viewing the primary sources, have to be performed leaving EHPS itself, for the web-site where the sources are directly viewable. Performing a search in EHPS is done using of four different ways [18].
The simplest one is to browse one of the five categories offered in the left-hand column: Country, Language, Period, Subject, Type of source.
The most selective one is to search the portal's combination of tags [19], the list of meta information divided within five different categories: a chronological one from the medieval times to nowadays; a linguistic one dealing with all the European languages in which the primary sources may have been written; a list of countries on which the sources are telling something as single nation's history or on Europe as a whole; a typology of sources trying to define different kind of available digital documents and, finally, a broad list of subjects under which the primary sources may belong to. A third primary source retrieval possibility is offered with a free text search -also in advanced mode- in the whole indexed content of EHPS [20]. This search is useful when names, places, titles of primary sources are already known by who's using EHPS in order to discover them through the descriptive in nEnglish abstracts created for each single web-site indexed.
A fourth way to search for contents is going also outside EHPS itself, using the Google Custom Search [21] for a wide search inside the listed web-sites indexed.

INTERACTING WITH EHPS CONTENTS

Being informed about new contents is easy receiving RSS Feeds. Following EHPS directly from Twitter is also possible. Each time a new web-site is added to the portal, whoever follows EHPS receives a tweet with the title of the indexed site and URL to visit it [22].
There's also the possibility to subscribe to become a fan of EHPS in the Facebook and to be informed on all new entries there too [23].
Registered users [24] have the possibility to "vote" for the qualities of the indexed web-site and comment, annotate and complete the abstract and the description of each single web-site with their own information. A registered user is also able to create ones own list of bookmarks and to suggest via a web-form, new web-sites to be included in the portal. Some News are also directly available on the portal itself [25].
Google Analytics is monitoring accesses to the portal. After few month of activity, we may say that some of the first web-sites indexed were already viewed more than 1.000 times each and the portal has been abstracted in Intute [26] and reviewed in H-Soz-u-Kult [27] and is connected within specific widgets in some library and digital humanities web-sites [28].

CONCLUSIONS

What would of course change enormously the importance of the portal [29] would be to become as complete as possible including all primary sources on the history of Europe completing the Multilingual inventory of Cultural Heritage in Europe portal and other European portals which are also included in EHPS [30] so to become the main history portal and reference for multi-lingual contents on the history of Europe.
If the consensus of the community of historians will grow, who knows ? Until now, accesses to the portal have increased exponentially during a year of activity [31] .

NOTES

[1] Marc Bloch: The historian's craft, New York, Vintage Books, 1953.
[2] Serge Noiret: "History in the EUI Library: A Retrospective", in EUI Review, Spring 2008, pp.30-31, <http://www.eui.eu/PUB/EUIReviewPDF/EUIreview_spring_2008_LIGHT.pdf>.
[3] Roy Rosenzweig: "Scarcity or Abundance? Preserving the Past in a Digital Era, in The American Historical Review, Vol.108, n.3, June 2003, <http://www.historycooperative.org/journals/ahr/108.3/rosenzweig.html>.
[4] Kirsten Sword in "Interchange: The Promise of Digital History", in The Journal of American History, 95.2, September 2008, § 35, <http://www.journalofamericanhistory.org/issues/952/interchange/index.html>
[5] Steven Mintz in Ibid, § 195..
[6] Digital Humanities Luxembourg: Contemporary history in the digital age, [http://www.digitalhumanities.lu/Pages/default.aspx>.
[7] CLEO, <http://cleo.cnrs.fr/>.
[8] Revues.org, <http://www.revues.org/>
[9] Noga Arikha and Gloria Origgi: Text-e, <http://www.text-e.org/index.cfm?switchLang=Eng>.
[10] Theodore Zeldin: An intimate history of humanity., London : Vintage, 1998.
[11] Serge Noiret, "Internet et les internet(s), lettre à Théodore Zeldin, Vendredi 14 Décembre 2001", in The Future of the Internet: A conversation with Theodore Zeldin, <http://www.text-e.org/debats/index.cfm?fa=view&type=view&ConfText_ID=9&Parent=0&Intervention_ID=332&Top_ID=332>
[12] Edward L Ayers: «The Academic Culture & The IT Culture: Their Effect on Teaching and Scholarship", in EDUCAUSE Review, v.39, n.6, pp.48-62, Nov-Dec 2004, p.55, <http://net.educause.edu/ir/library/pdf/ERM0462.pdf>.
[13] Daniel J. Cohen and Roy Rosenzweig: Digital history: a guide to gathering, preserving, and presenting the past on the Web., Philadelphia: University of Pennsylvania Press, 2005.
[14] European University Institute, <http://www.eui.eu>.
[15] WWW VL History Central Catalogue, <http://vlib.iue.it>.
[16] Dries Buytaert, <http://buytaert.net/resume> and Drupal, <http://drupal.org/>.
[17] Zen, <http://drupal.org/project/zen> and John Albin Wilkins, <http://www.albin.net/>.
[18] How to search the portal, <http://primary-sources.eui.eu/search>.
[19] Category browser, <http://primary-sources.eui.eu/combined-category-search>.
[20] Free Text Search, <http://primary-sources.eui.eu/free-text-search>.
[21] Google Custom Search Engine, <http://www.google.com/cse/>.
[22] EHPS Twitter account, <http://twitter.com/EHPS>.
[23] EHPS in Facebook, <http://www.facebook.com/pages/European-History-Primary-Sources/223099761969>.
[24] User account. Create a new account, <http://primary-sources.eui.eu/user/register>.
[25] News Archive, <http://primary-sources.eui.eu/news/>.
[26] Intute, <http://www.intute.ac.uk/cgi-bin/fullrecord.pl?handle=20090121-13211595#user>.
[27] Martin Munke: Web-Rezension zu European History Primary Sources EHPS, in: H-Soz-u-Kult, 07.11.2009, <http://hsozkult.geschichte.hu-berlin.de/rezensionen/id=158&type=rezwww>.
[28] Repertorio di risorse web a cura della Biblioteca di Filosofia e storia dell'Università di Pisa, <http://filosofiastoria.wordpress.com/>;
[29] MICHAEL, <http://www.michael-culture.org>.
[30] Other Portals, <http://primary-sources.eui.eu/other-portals>.
[31] EHPS was officially launched the 8th of June 2009. Alexa Traffic rank for EHPS, <http://www.alexa.com/siteinfo/http%3A%2F%2Fprimary-sources.eui.eu#trafficstats>, 8th December 2009.

Abstract
Digitisation provided enormous opportunities for presenting cultural heritage online. Archives, libraries and museums are commonly entitled as memory institutions and digitisation is considered a useful tool for performing their memory function. Currently, there is a tendency of aggregation of heritage content by building international digital portals oriented at the general public. However, many questions about the nature of memory and its communication, benefits gained by general public using the online collections etc. remain unexplored. The aim of this paper is to explain how memory is communicated in archives, libraries and museums basing on the theories of memory and heritage, and to demonstrate by employing the examples of digitisation initiatives how these theories can be applied for developing online heritage services. As a result of theoretical analysis and case studies a conclusion that digital conversion and online access should not be an ultimate goal of the initiative was made. When developing the concept of digitisation project memory institutions should bridge the idea of heritage service with the current context and needs of the user community. Technologies reflecting common memory practices (e.g. web 2.0 photo sharing systems) and providing attractive communication environment have a great potential for developing online heritage services.
Keywords: digitisation, memory, memory institutions, online heritage services, web 2.0

Introduction
The term "memory institution" originates from the recognition of the significance of memory function which is one of the essential pre-requisites for the existence of archives, libraries and museums. Digitisation offered new opportunities to communicate memory; however, it also brought the challenges of discovering meaningful ways of such communication in the changing socio-cultural and technological environment.
The major advantage offered by digitisation is an opportunity for wide access that allows overcoming geographical and time barriers. Inspired by access potential for cultural heritage archives, libraries and museums engaged in digitisation projects that evolved rapidly from small one institution experiments to large-scale initiatives implemented by international consortia. Today there is a time of large-scale online cultural heritage services as the *European Library*, *Europeana*, the *World Digital Library* etc. Majority of these international initiatives pursue the objective of providing rich collections to diverse audiences like general public, scholars, educators and others.
Such concepts as memory, heritage, knowledge are often put together in diverse visionary statements. "Information technologies can enable you to tap into Europe's collective memory with a click of your mouse", as stated by the Information Society & Media Commissioner Viviane Redding in her speech [1]. Later on, in the European Commission communication on the progress of digitisation, accessibility and digital preservation of cultural heritage in Europe the metaphor "collective memory with a click of your mouse" transformed into "Europe's cultural heritage at the click of a mouse" [2].
The visions formulated by politicians and large-scale digitisation consortia raise a lot of questions. How the general public benefit from cultural heritage collections? How digitisation relates to memory? Are cultural heritage and memory communicated in the same way? Is it enough just to put digitisied materials online? Current research reveals that many contemporary digitisation initiatives ignore the needs and the nature of establishing links with the past by individuals and communities. Research of online cultural heritage initiatives performed by the American researcher Maria Dalbello has shown that in many cases project ideas were based on cultural heritage collection – its structure and content [3], but not related to the needs and expectations of its potential users. Similarly, the analysis of cultural heritage projects supported by the European Union perfomed by Zinaida Manžuch revealed that project initiators were more interested in issues of managing cultural heritage resources than constructing meaningful stories of the past [4].
The aim of this paper is to explain how memory is communicated in archives, libraries and museums basing on the

Chiara Cirinnà and Maurizio Lunghi (edited by), *Cultural Heritage on line: Empowering users: an active role for user communities. Florence, 15th - 16th December 2009* ISBN 978-88-6453-184-7 (print) ISBN 978-88-6453-187-8 (online)

theories of memory and heritage, and exhibit how these theories can be used for building cultural heritage services in the digital environment.

Theory: why and how the modern societies remember

Memory is a way of individuals and societies to deal with the past. The word „re-membering" means becoming a member again and indicates that memory is a source of social cohesion in human communities [5]. Although by re-membering we create the link to the past, it is not re-living it again. It is impossible to re-live again the same emotions or enter the same contexts or events as they were in the past. By remembering the past is always constructed again according to the present condition, views and needs of those who remember [5]. Individual recollections are influenced by the membership in communities that form the social memory environment. Communities are remarkable for common needs and interests, which become what Maurice Halbwachs called 'les cadres sociaux de la mémoire' [social frameworks of memory]. Individuals "recall" events or experiences that may precede their birth, and these recollections are very similar within the same communities. Social frameworks, in Halbwachs words, are '… precisely the instruments used by the collective memory to construct an image of the past which is in accord, in each epoch, with the predominant thoughts of the society' [6: p. 40].

Cultural heritage may be approached a mnemonic device that connects us to the past. When there are no living witnesses to tell the story of the past, events or experiences transform into remote symbols and rituals that become a part of the identity and history of a particular community. These stories of the past are mediated by cultural heritage [7].

Cultural heritage is a part of the past selected in accordance with contemporary needs of societies [8]. It is loaded with constantly shifting symbolic meanings. The meanings of heritage are different from memory per se, in that they are explicit: they may be disseminated and comprehended by any person. Memory always means "being" or "belonging", while heritage may also mean "knowing about", which is open for rational cognition. Only common life context, values and experiences enable the transformation of heritage into memory. In other cases, heritage may be valued for particular features, used in education, which have nothing in common with memory. Therefore, interpretations of heritage symbolic meanings also allow archives, libraries and museums to communicate knowledge about the past without any reference to the collective memory of communities [4].

In the modern societies remains of the past are transformed into heritage as a result of selection decisions performed by different institutions that usually include research organisations, memory institutions, governmental bodies etc. In archives, libraries and museums, for instance, particular items become cultural heritage as a consequence archival appraisal or well-defined selection processes in libraries and museums. The process relies on expert views on societies, their needs, contemporary context and values. This inevitably creates a gap between heritage experts and citizens. Therefore, cultural heritage does not become a heritage in a full sense until it is acknowledged by the society [9]. When cultural heritage is meaningful to citizens they are able to remember, i.e. construct their images of the past. The major role of memory institutions is to interpret and contextualize cultural heritage for it to become meaningful to people in their present lives.

Communication of memory in archives, libraries and museums is also shaped by the development of new media that introduces novel ways of interaction with cultural heritage collections. The impact of the communication technologies (in a broad sense – language, print, digital media) to the ways a human being understands the surrounding world and shares these meanings with others was widely researched and argumented. Contemporary internet technologies (web 2.0 in particular) are remarkable for growing user-centeredness, focus on co-authorship and collaboration, interactivity of services [10]. Although all ideas about computer and web are rooted in older communication technologies, when used they undergo enormous transformations. This argument was illustrated by Christine Borgman reflection on email which was rooted in metaphors of traditional paper post and letters but evolved into an interactive service with newly developed language (i.e. short messages and abbreviations) and elements (e.g. smileys) [11]. In order to meet user requirements memory institutions should adapt to evolving communication technologies, environments and practices.

Practice: communicating memory in digitisation initiatives

The essential feature of memory and heritage is the link between the past and present context of community life. Memory institutions should consider these links while formulating the objectives of digitisation initiatives. In this paper

the focus is put on services related to communication of memory; thus excluding all spectrum of services aimed at communication of heritage (e.g. services for scholars, learners, cultural tourists etc.).
Archives, libraries and museums often employ two ways of linking the past with the present. The first way involves orientation at the identity, values and structure of particular community, while the second – particular social issue or phenomenon that creates new communities. The instance of the first way are geographical communities, who mainly associate their identity and personal history with particular place. The instance of geographically oriented digitisation initiative is the project Worthington Memory (USA) undertaken by the Worthington libraries and Worthington Historical Society. The initiative was grounded on the needs of the local community to cultivate social bonds and determine identity of the community [12]. One of the remarkable opportunities offered to the users is so-called Time Machine, the service enabling the user to compare past and present views of particular geographical objects by manually uncovering the 'elderly' image layer placed under the modern photo of the item. The photos showing changes of the city become meaningful for local citizens, even those, who have never seen before how the object looked in the past.
Communicating memory by focusing on a specific social issue/phenomenon is illustrated by the project Moving Here (UK), which was implemented by the consortium of 30 memory institutions and dedicated to exploration of 200 history of immigration to the country [13]. With increased migration flows this issue of the past is very actual to both those who immigrate to UK nowadays and those who face the challenge of multicultural society. Each person, visiting the portal can write a digital story about his/her own migration experiences. Sharing personal migration stories bridges past events with the current experiences and thus makes cultural heritage collections meaningful mnemonic devices for the contemporary migrants.
Examples of adoption of new environments for communicating memory is illustrated by the growing number of heritage initiatives that employ web 2.0 tools. Two collaborative initatives with Flickr – PictureAustralia (Australia) and Library of Congress pilot project (USA) – are relevant instances of successful web 2.0 application for the development of heritage services. In the first case, the national online historical image service PictureAustralia was enhanced by enabling users to contribute their own photos. The service became available due to the partnership with Flickr that provided attractive and easy-to-use interface for uploading and describing the photos. One of the outcomes of the initiative was the increased participation of user community, who actively engaged in uploading personal photos and re-photographing historical images. The participative service also increased the visibility of PictureAustralia portal and its usage [14]. Another instance of successful web 2.0 uptake is the initiative of the Library of Congress (USA) aimed at representing historical photographs on Flickr [15]. The project exhibited new ways of engaging communities in cultural heritage interpretation, i.e. remembering: "It is particularly gratifying to see Flickr members provide all kinds of connections between the past and the present through discussions of personal histories including memories of farming practices, grandparents' lives, women's roles in World War II, and the changing landscape of local neighborhoods" [15: p. 26]. The project evolved into large-scale international consortium Flickr: the commons, which covered cultural heritage institutions from various countries.

CONCLUSIONS

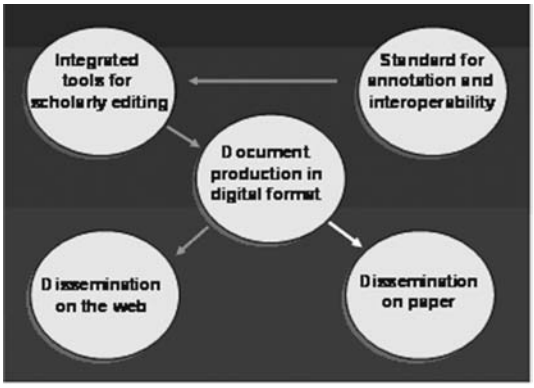
Memory institutions, initiating digitisation projects, should realise that digital conversion and online access are not the ultimate goals of the project, but rather tools for developing heritage services. Understanding of cultural and social mechanisms of remembering should guide the development of service concept. In most cases representative of general public have no definite information need but is willing to reinforce the feeling of belonging to particular community. Examples illustrated that the links between heritage and communities could be established by making references to important geographical coordinates of community life and contemporary social issues that were also faced in the past. Heritage is not memory; therefore, the definitive feature of a memory institution is not merely holding a cultural heritage collection, but also performing activities that transform heritage into the cultural intermediary of memory. Heritage becomes meaningful mnemonic device when it is related to the present life context of communities and individuals. It is the responsibility of memory institutions to make the links between the past and the present meaningful to the user.
Experiences of current digitisation initiatives have shown that reliance on widespread memory practices and popular communication tools is an effective way of engaging users in recollection and active interaction with cultural heritage collections. Successful experiments with Flickr proved that this web 2.0 tool utilising popular habits to collect photo

albums and imitating common actions of making descriptive notes allows developing attractive environment for communicating memory.

REFERENCES

[1] "European Commission steps up efforts to put Europe's memory on the web via a "European Digital Library" (2006), available at: http://europa.eu/rapid/pressReleasesAction.do?reference=IP/06/253 (accessed at 15 October 2009).
[2] "Europe's cultural heritage at the click of a mouse: progress on the digitisation and online accessibility of cultural material and digital preservation across the EU"(2008), available at: http://ec.europa.eu/information_society/activities/digital_libraries/doc/communications/progress/communication_en.pdf (accessed at 15 October 2009).
[3] Dalbello, M. (2004), "Institutional shaping of cultural memory: digital library as environment for textual transmission"Library Quarterly, vol. 3, issue 74, p. 265-299.
[4] Manžuch, Z. (2009), "Archives, libraries and museums as communicators of memory in the European Union projects", Information Research, vol. 14, issue 2, available at: http://informationr.net/ir/14-2/paper400.html (accessed 15 October 2009).
[5] Olick, J. K., Robbins, J. (1998), "Social memory studies: from "collective memory" to the historical sociology of mnemonic practices", Annual Review of Sociology, vol. 24, p. 105-140.
[6] Halbwachs, M. (1992). On collective memory. Chicago, IL: University of Chicago Press.
[7] Assmann, J. (2004). Kul'turnaya pamiat': pis'mo, pamiat' o proshlom i politicheskaya identichnost' v vysokih kul'turah drevnosti. [Cultural memory: literacy, memory of the past and political identity in the high cultures of antiquity.] Moskva: Yazyki slavianskoy kul'tury.
[8] Graham, B., Ashworth, G.J., & Tunbridge, J.E. (2004). A geography of heritage: power, culture and economy. London: Arnold.
[9] Čepaitienė, R. (2005), Laikas ir akmenys: kultūros paveldo sampratos modernioje Lietuvoje. Vilnius: Lietuvos istorijos instituto leidykla.
[10] Musser, J., O'Reilly, T. (2007), Web 2.0 principles and best practices. Sebastopol, 2007.
[11] Borgman, C. (2000), From Gutenberg to the global information infrastructure: access to information in the networked world. Cambridge: MIT Press.
[12] "Worthington memory"(2009), available at: http://www.worthingtonmemory.org/ (accessed 15 October 2009).
[13] "Moving Here – 200 years of migration to England" (2009), available at: http://www.movinghere.org.uk/ (accessed 15 October 2009).
[14] Hooton, F. (2006), "Democratising history: evaluating PictureAustralia's Flickr pilot project", National Library of Australia Gateways, December, no. 84, available at: http://www.nla.gov.au/pub/gateways/issues/84/story13.html (accessed 15 October 2009).
[15] Springer M. et al. (2008), "For the common good: the Library of Congress Flicrk pilot project", available at: http://www.loc.gov/rr/print/flickr_report_final.pdf (accessed 15 October 2009).

Over the past years text processing systems have become part of the daily language of many scholars working in the field of the Humanities, despite some objections raised against this type of technology which still seems to be distant in terms of simplicity of usage, appropriateness, and flexibility.
Usage requires particular attention as concerns the interface between the information system and the user, while appropriateness and flexibility have not been sufficiently taken into consideration even due to the fact that they almost seem to be in contradiction. Therefore, it is not easy to plan and implement a text processing system which is suitable for specific types of research and at the same time flexible to operate in various sectors of study.
The project Pinakes Text (PKT) that I am presenting here, is aimed at achieving this ambitious target through an architecture based on interconnected modules. In other words, the system works with a nucleus of components for the treatment of both text files and digital image files, which form the core of the system. According to the specific needs, from time to time a number of programs are added both for the management of images (enhancement, segmentation, pattern recognition, etc.) and of text (natural language processing, information extraction, data mining, electronic editing, ecc.). The simplified scheme which lies behind PKT could be represented as follows:



1. The first element is represented by the respect of internationally shared standards, so that the information managed by PKT is interoperable with other data produced in the humanistic field. The standards are also followed when not only primary data (texts, images, etc.), but also secondary information, such as annotations, variants, comments and/or information produced by computational systems (e.g. morphological, syntactic, semantic analyses) are introduced. The software development tools are totally open source in order to avoid any fees for end-user licences.
2. The information system is entirely web-based and the tools for the production or search for information are oriented towards the sector of critical and textual scientific editing. At present, the target of PKT is represented by specialist users. However, the structure of the system also envisions a number of operations, in particular those connected to the phases of search and query, which can be further developed so as to meet the needs of a non-specialist-user.
3. PKT allows to produce on a web server data that have been labelled and annotated in collaborative form, as long as all the members of the same community (e.g. mediaeval philologists, Greek papyrologists, Egyptologists, Latin epigraphists, historians and science philosophers, etc.) agree with the same standards, as evidenced in point 1.

Chiara Cirinnà and Maurizio Lunghi (edited by), *Cultural Heritage on line: Empowering users: an active role for user communities. Florence, 15th - 16th December 2009* ISBN 978-88-6453-184-7 (print) ISBN 978-88-6453-187-8 (online)

4. Some experiments are in course to check whether PKT meets the needs requested by a community of scholars working on documents of Egyptian archaeology. The documents and annotations are produced in digital format and are classified according to a domain ontology agreed upon by the same members of the community. This semantic-conceptual structure can be replicated not only to classify the documents, but also part of their content. In this way, it is possible to retrieve information both at the level of forms (words, strings of characters, lemmas), and at the level of concepts expressed in the single parts of the texts.

The main form of dissemination envisioned by PKT is the one on-line on a server permanently connected to the internet. The encoding of the data, entirely performed in XML language, also allows distribution in paper format. As concerns this particular aspect, for the next phases of development of the system, the modes of production of the information managed by PKT on e-book will be taken into consideration. The introduction of e-book on the market should provide considerable medium-term increase percentages.
With regard to the current use of PKT:

- it is one of the text and image management components participating in a COST Action of the European Science Foundation (http://www.interedition.eu/);
- it has been considered the suitable technological basis for the project ERC (Advanced Grant) "Greek into Arabic", approved and financed in the first days of November 2009 by the International SH5 Panel of ERC;
- it manages the corpus of the National Edition of the works of Galileo Galilei (http://pinakes.imss.fi.it:8080/pinakestext/home.jsf);
- two European project proposals are underway which in case of success could consolidate the position of PKT as infrastructure of research in the sector of sciences of the text.

Implementation of the system is the result of a collaboration between the Institute for Computational Linguistics "A. Zampolli" of the National Research Council of Pisa, the "Fondazione Rinascimento Digitale" and the Institute and Museum of the History of Science in Florence.

Abstract
This paper explores the pedagogical concepts and technological framework underlying the MyDante Project (http://dante.georgetown.edu), developed by the Center for New Designs in Learning and Scholarship at Georgetown University. We argue that MyDante, essentially a Web-based, interactive hypertext edition of Dante's Divina Commedia, represents a distinctive approach to designing spaces in which readers interact with texts. Always acting on the principle that the technological design should serve the pedagogical goals of the project, we have created an environment to facilitate deep engagement with Dante's poem within the context of a particular method of study known as contemplative reading. MyDante encourages the reader to experience the poem in a way that is profoundly personal, while at the same time enabling a collaborative experience of the shared journey by a community of readers, and creates a record of that journey.
After exploring the interplay between the pedagogical aims and technological structure of MyDante, we briefly describe the current development of Ellipsis, a new iteration of the application which will be customizable for a wide range of texts, disciplines, and pedagogical approaches. Finally, we provide some examples of how students experience MyDante, and conclude with further reflection on the distinctiveness of MyDante's approach.
Keywords: MyDante, Digital Humanities, pedagogy, technology, contemplation

Introduction

If we think of any reading experience as a meeting of a text and its context with a reader and his or her context, then we might say that those of us in the Digital Humanities have tended to focus much of our energy on understanding, marking up, and making available the first half of this equation. That is, the field of Digital Humanities writ large has often been defined by textual projects that emphasize the importance of the texts and their contexts (alternative versions, supplemental texts, historical documents, and so on). The past twenty-plus years of research projects within the Digital Humanities have made available to the humanities scholar a wide range of historical, literary, and cultural texts, marked up for quick searching, complex linguistic analysis, and archiving. We believe that, in many respects, the MyDante project (http://dante.georgetown.edu) represents an alternative approach to the Digital Humanities, to our focus on the text and its primacy, and to our understanding of the relationship between the reader and the text.
The MyDante project began a little over ten years ago with the primary goal of providing students with a contemplative space in which to engage with Dante Alighieri's Divina Commedia, specifically within the context of the undergraduate philosophy class "Dante and the Christian Imagination" at Georgetown University. Many of the previously available online versions of the Commedia focus on markup, searching, and archiving the text. From the beginning, the MyDante project was designed to enable students to understand the text through their interaction with it, their reflection on it, and their engagement with their peers around it. Inspired by the metaphor of the medieval illuminated manuscript, we aimed to allow students to see the text of Dante's poem as a palimpsest, as a place where their ideas and their writing share the same space as the poem; where they could engage with and rethink the poem just as a monk in the Middle Ages might have done through marginalia and illuminations. We created and continue to develop a variety of tools, such as an annotation tool, a journaling tool, and a multimedia editor, to encourage students to interact with the poem and share their ideas with others, much in the way the marginalia of a medieval manuscript would influence future readers. MyDante simultaneously encourages deeply personal reflection as well as scholarly collaboration focused on the text. Dante's poem is particularly well suited to this type of contemplative experience, but we see this process being extended to virtually any text. MyDante gives us a powerful model that can be localized to the classroom or extended to larger communities interested in reading interactively and reflectively. To facilitate this type of reading experience within a wide range of contexts, we have created a digital publishing platform we call Ellipsis (see Section 2). In many ways,

Chiara Cirinnà and Maurizio Lunghi (edited by), *Cultural Heritage on line: Empowering users: an active role for user communities. Florence, 15th - 16th December 2009* ISBN 978-88-6453-184-7 (print) ISBN 978-88-6453-187-8 (online)

the types of tools we attempted to create ten years ago prefigured much of what has become standard practice on the Web as we've moved from a read-only Web to a read/write Web. Social networking tools such as Diigo, YouTube, and Wikipedia encourage all Web users to see the Internet as a shared text to be viewed reflectively and even at times contemplatively.

The Pedagogy Underlying MyDante

The goal of the MyDante project is to make available a Web-based, interactive hypertext edition of Dante's epic poem that allows each reader to develop an illustrated, annotated, personalized copy of the text over the course of a lifetime. More than just a technology, MyDante offers a pedagogy of reading and reflection designed to illuminate and document the reader's experience of the poem. Dante wrote the Commedia as an invitation to undertake a journey of self-discovery. MyDante serves as a permanent record of accepting that invitation and sharing Dante's journey. It makes the poem profoundly personal, while at the same time enabling a collaborative experience of the shared journey by a community of readers.

A fundamental principle of the project from the beginning has been that technology by itself, no matter how engaging, is only an instrument and must be given a humane purpose. The pedagogy that informs the MyDante project, inspired in part by Dante's "Letter to Can Grande della Scala," is essentially the method of allegorical interpretation developed by the Jewish and Christian traditions to read the Scripture faithfully and which was perfected by the contemplative culture of medieval monasticism. All aspects of the medieval monastic culture were pedagogically integrated around its central purpose and activity: contemplation. The monks developed practices of active engagement with Scriptural texts, from meticulous manual copying to artistic illustration and scholarly commentary. From the manuscripts the monks produced, we can clearly discern how closely connected technology, action, and contemplation are in the fabric of a human way of life.

- What must a reader do to experience Dante's poem contemplatively? How does reading as a contemplative practice work? The first requirement is to recognize that, just as Dante told his patron Can Grande della Scala, there are multiple levels of meaning simultaneously at work in the poem, each of which necessitates a different kind of understanding; the second requirement is to learn how to move progressively from:
- The literal, narrative level of the story of Dante the pilgrim's journey from the Dark Wood to the Final Vision. The goal of reading at this first level is clear comprehension of the characters and plot. From here to
- The ironic and metaphoric level communicated by the artistic choices Dante the poet makes regarding characters, episodes, images, and themes. The goal is to arrive at an interpretation of the poet's message to the reader; that is, to explore the questions, Who is Dante the poet? What is he trying so hard to tell me? Then, from this level to go to
- The reflective level, contained in the reader's personal responses to the poet's confession, witness, and testimony, in the form of a dialogue between poet and reader. The goal of this level of understanding is personal reflection. The meaning of the poem is not finally understood until reader and poet find themselves standing face to face, in the presence of all others who confront the same questions of personal identity, freedom, and responsibility. To do this requires the reader to imagine how one's own journey is the same as the poet's, how both are "universal," the same for all persons, despite every difference of time, place, and culture. To understand the poem, then, is to become part of the poem by recognizing oneself in it and by making it genuinely one's own by responding to the question "Who am I?" - not simply as an individual, but as a person who is both the same as and singularly different from every other person in such a way that, as the poem tells us from the beginning, the story it recounts is truly the story of "our life."

Readers can move among these three levels of understanding by using MyDante's tools, which include:

- Side by side Italian and English texts and an Italian audio recording of the poem
- General introductions to each of the Cantiche, including some in video format
- An Image Gallery, containing illustrations of the Commedia and a wide range of images with thematic connections; users can also upload their own images.
- An annotation tool, mirroring the function of manuscript marginalia, allowing the reader to comment on specific lines and read others' comments

- A journaling tool that enables readers to embed extended reflections into MyDante, creating personal records of their relationship with the poem
- A Biblioteca, which houses digital copies of other works by Dante and relevant texts by other authors, as well as chronologies, maps, and other resources
- A Chapter Room that includes a space for interactive discussion

Given the broad pedagogical goals outlined in the three levels of reading above, in what follows we shall focus on a specific question we faced in developing MyDante: to what extent can digital media effectively guide readers in comprehending, interpreting, and reflectively appropriating the significance of texts, either as an enhancement of the role of human teacher or, in certain cases, as an alternative to direct contact with a personal guide?
First, one clarification: this question is not properly subsumed under the heading of "distance learning," or delivering content and evaluating student mastery of materials or skills online. In the context of Ellipsis/MyDante, the scenarios we envision are either the enhancement of live teacher effectiveness through the transfer of certain components from the classroom to students' independent work outside class - so as to free classroom time for different and arguably more advanced learning activities - or, on the other hand, the creation of a community of readers who are not enrolled in any formal academic program and who are not
receiving direct personalized guidance or evaluation from an instructor.
In other words, we asked ourselves how much learning at the level of direct engagement with literary texts could be achieved via a pedagogy that subsists solely within the structure of a digital platform, and to what extent this platform could open up new possibilities for further learning. The challenge we faced was how to deliver digitally not just a broad array of diverse content material, however well organized and intuitively accessible, but to deliver such content in the context of an artfully designed pedagogy that provides the student/reader with skilled guidance that would otherwise be either absent or less efficiently and effectively delivered in person.
One of the most challenging questions we addressed was how to activate and deliver the pedagogical design that would guide students through the threefold dynamic of comprehension, interpretation, and reflective appropriation outlined above. The solution was to develop a Guide function that could overlay the text and be turned on or off by the reader. More than a commentary or scholarly investigation, the Guide would directly engage students in a series of activities, such as listening to the poem in Italian, dwelling on the details of an illustrative work of art, reading commentary on the metaphoric range of a particular symbol or theme, writing a journal entry reflecting on the larger human significance, for example, of Dante's placement of Ulysses in Hell—activities that would both model learning behaviors for students and stimulate them to repeat and develop those behaviors for themselves.
Both technically and pedagogically, developing an effective Guide format proved even more of a challenge than the team had anticipated and required various stages of experimentation, implementation, evaluation, and revision over the course of two years and two iterations of the class "Dante and the Christian Imagination" at Georgetown. The highest hurdle was how to get students started using the site and to become quickly at home with MyDante's pedagogical tools and practices. The solution we are now testing is an introductory "Getting Started" Guide, designed to supplement class lectures and demonstrations. By combining video segments, audio commentary, screen-capture tutorials, thematic commentary, and a reflection on contemplative reading, we believe we have taken a substantial step toward giving students access to not only a rich body of content but also a progressive, self-regulating pedagogy with which to approach this vast literary monument of human culture and even to draw on for future study of other texts.

DESIGNING TECHNOLOGY TO SUPPORT PEDAGOGY

In order to serve the pedagogy outlined above, the MyDante application needed to enable and encourage certain types of activities by the users. First, the site had to direct users inward, toward a deepening contemplation of the primary text, even as it facilitated comparisons among a broadening scope of related texts (in various media - documents, images, sounds, videos, and even animations). Second, MyDante needed to engage readers with one another as a community of scholars, encouraging them to explore one another's emerging perspectives as a way of continually rediscovering their own interpretations of the poem's significance.
Consideration of the philosophy behind the pedagogy played no small role in developing the application's functional

design. Rather than focusing on creating a sophisticated markup to encode a particular domain of inquiry around the poem, or on providing an entirely open-ended discussion forum, we chose to integrate several simple actions, which we felt best supported the invitation to students to deepen their personal involvement with the text, within the context of a coherent pedagogical framework. This choice reflects a conviction that readers are ultimately the repository of the deepest meanings in a text. It is the interpersonal connections among a community of readers, moved by a text in similar and different ways, that ensure its survival as a work of art.

The Guide function described above, although by no means the only tool designed to invite students into a deeper reading of the text, allows us to trace how the application supported each level of reading. The Guides address the first level of comprehension by providing contextual historical and literary information in the form of brief annotations similar to footnotes. At the second (metaphoric) level of reading, Guide materials interspersed throughout the poem encourage specific readings of particular themes and symbols, and point out resonances among different sections of the text. Finally, at the third (reflective) level, the Guide asks open-ended questions which prompt directed reflection by the students on how the poem might relate to their own lives.

Our intention was to provide a self-contained digital cloister within the chaos of choices available on the Web. We discovered that by creating ways to link parts of documents with one another, documents that comprise a system of meanings according to a clear pedagogy, we were able to channel the energy of browsing into a directed activity of probing. We realized that this capacity for focusing a reading community's activity around and concentrically toward the meanings of a text would be very powerful in a wide variety of educational contexts. To that end, we have abstracted the functionality developed for MyDante into an application framework we call Ellipsis. Ellipsis promises to provide, for a wide variety of texts in a wide variety of media, a sharable space for study and reflection. The basic tasks of making a statement about a text or some selection within it, and relating it to some other text (or selection of text), whether from primary materials or those generated within the community, are abstracted to apply similarly to images, sounds, video and other media. In addition, collaboration between faculty is encouraged by allowing multiple overlapping spaces, where community members can easily move into and out of collaborative exercises with members of other communities centered around the same system of related texts.

Evidence of Student Learning

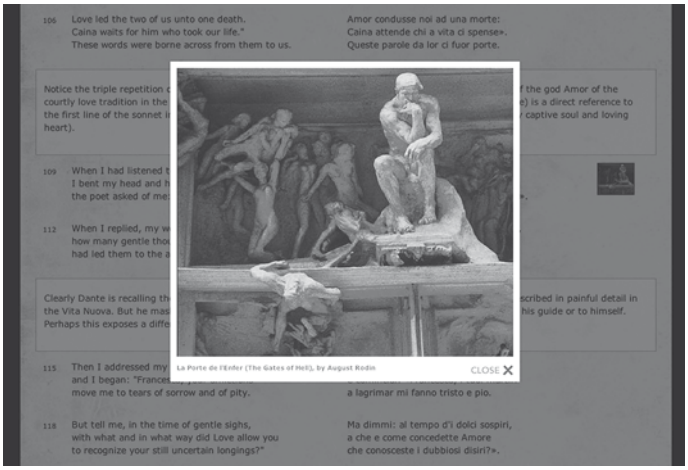
Throughout the project, we have collected feedback and suggestions from students through surveys and interviews, and we have gathered evidence of student learning from their work on the site. Student response to MyDante has generally been extremely positive. For example, one student explained that "it added a different dimension to the class that there was so much that you could experience outside of it. It was a very holistic learning experience, and it added a whole new level of discussion - it was like having twice as much class time." Another student finds himself drawing on MyDante's approach long after completing the course: "MyDante's comprehensive resources and scholarship will always be something I turn to, and the site is an ideal model for both how to read such a vast work and how to contemplate its richness... My experience with MyDante is invaluable in continuing to interpret the poem's meaning and in forming strategies for evaluating other texts."

After we implemented the Getting Started Guide in the fall of 2009, it became clear that the students were demonstrating a better understanding of the material than students had at analogous stages of earlier semesters. They had digested and integrated the various components of the Guide and were relatively comfortable with both the technological and theoretical practices involved in using the site effectively. This deftness was evidenced both by their comments in class and their contributions to the site, in the form of discussion board posts, annotations, and journal entries, in which they applied the concept of the three levels of reading to their interpretations of the text. The students also showed a high degree of engagement with one another on the site and in class. Many of them spoke in class in addition to contributing to the site, and many of them responded to their classmates' posts on the site. For example, after only four class sessions, a student's 1250-word post entitled "Why the Catholic Cosmos?" elicited more than 1700 words of comments by five classmates.

As the site evolves, we are continuing to evaluate student responses to various features of the site and to gather evidence of student learning. For examples of student work demonstrating particular pedagogical goals, and for more testimonials from students about their experiences with the site, please see http://dante.georgetown.edu/student_learning.

CONCLUSION

We see the MyDante project as representing a genuinely new approach to designing a space for interaction with texts. This approach enables a distinctively individual reading experience characterized by depth, richness, and intensity, an experience which is further enhanced by the collaborative dimension of reading within a virtual community. MyDante memorializes the reader's journey in a way that is simultaneously imaginative and technological in character, so that each reader's record acts as a personalized archive of meaning. As we continue to develop MyDante and Ellipsis, we remain committed to the principle that technology is subordinate to pedagogy in our aim to transform and enrich the ways in which readers engage with texts of all kinds.



Screenshot of MyDante (http://dante.georgetown.edu) showing the image viewer tool. Note the guide sections interspersed within the poem text.

Abstract
This paper reports on a study of the impact of, and challenges posed by information on the museum, and the changing nature of museum work. The study involved semi-structured telephone interviews with sixteen senior North American museum professionals and academics teaching in museum studies programs. Our findings suggest that the ways museums interact with their publics and the areas of dissemination and collections management are profoundly changing. It found that the three most common challenges that museums face include: the cost of designing, implementing, and maintaining technology; a lack of in-house expertise; and information management. The study also indicates that the museum profession is facing a generational shift and that younger professionals perceive technology as a ubiquitous part of their environment.
Keywords: Museums, information technology, museum workers

Introduction
Many suggest that museums have been transformed by their societal context and the proliferation of information technology (IT) in our contemporary moment. Parry argues that digitality "helped to support a realignment of museography that was taking place, from object-centred to experience-centred design." [1] Bearman and Gerber posit that new technologies have fundamentally changed the ways museums communicate. They state that, "since the late 1980s, computer-based interactive programs have delivered more varied and exciting information on the museum floor than traditional mechanical interactives or static signage. Today, a museum without a collections database and a Web presence is hardly considered professional." [2] However, they go on to note "not all institutions are using online access equally well." [2] The will of museums to participate may not be sufficient impetus and Loran found in a study of British National Museums that a favourable 'political context' energized museum uses of web services to reach broader audiences. [5]The 2002 DigiCULT Report analyzed the current state of technological deployment in the cultural heritage and indicated how a range of technologies could be used to unlock the potential of such cultural heritage institutions as museums [3].
There is a large and growing literature evaluating the intersection of new technologies and museums by writers such as Paul Marty, Ross Parry, Sarah Kenderdine, Fiona Cameron, and Katherine Burton Jones. This literature addresses issues ranging from museum informatics [as Paul Marty defines the field, "the study of the sociotechnical interactions that take place at the intersection of people, information, and technology in museums" [6]], to resource development (such as imaging, digitization, and integrated information systems), from new media technologies and critical digital theory (pertaining to object morphologies, virtual systems, digital objects, communication technologies) to visitor interaction and online technology. The diversity and profusion of this literature suggest museum professionals and funding agencies need a better understanding of the challenges information technologies pose for museums, and the ways museums are changes to meet these challenges. The technology observatory research of The DigiCULT Forum is indicative of the needs of the community for access to knowledge about emerging technologies [8] and how these needs might be met.

Methodology
To gain this understanding, we undertook a research project, funded by the Canadian Heritage Information Network (CHIN) , involving semi-structured telephone interviews . We devised a series of questions, grouped into 4 thematic sections: i. new technologies used in the participant's own institution; ii. new technologies and the museum in general; iii. IT skills and training in museums; and iv. information technology and museums in the future. Telephone interviews took place between March 24 and April 9, 2009 and were conducted in either English or French, as appropriate. In consultation with CHIN, we identified twenty potential senior North American museum professionals and academics, whom CHIN contacted and invited to take part in the study. In the end due to time constraints on the part of potential interviewees we interviewed only sixteen individuals. The participants included three directors of collection management departments, three museum directors, three directors/chiefs of museum technology departments, three academics teaching in museums programs, one director of a museum education department, one project manager, one

Chiara Cirinnà and Maurizio Lunghi (edited by), *Cultural Heritage on line: Empowering users: an active role for user communities. Florence, 15th - 16th December 2009* ISBN 978-88-6453-184-7 (print) ISBN 978-88-6453-187-8 (online)

curator and one head of a museum standards program. We sent documentation to each participant in advance of their interview including a description of the project, a consent form, and an interview script. The researchers conducted the 30 to 75 minutes interviews, and digitally recorded all but one of the interviews. Each researcher took notes during her interviews, and subsequently replayed the recordings to gather additional comments and ideas. After completing all interviews, the research group met and reviewed the interview notes question-by-question in order to share findings. The research team identified recurring themes represented in the interview data and noted differing points of views. The research team then identified examples that illustrated these themes.

THE FINDINGS

The Impact of New Technologies on Museums

Interviews revealed that there is no consensus on the extent to which new technologies have impacted museums. One interviewee argued that the museum has not been transformed by new technologies any more than society in general. That is, technology has not altered the traditional core missions of museums to collect, preserve, interpret, and make available cultural heritage. Nevertheless, it has allowed museums to express their missions to a broader audience, and to fulfill them in a variety of new ways. Another interviewee suggested that, although technology has not changed museums' rapports with the public, it has changed the way museums work. Others suggest that technology has changed the way museums think about themselves. For example, one interviewee suggested that information technology has changed the way her museum views itself. She noted that the days of circulating collections are over but the museum can now provide web seminars to teachers in remote communities or other parts of the province. Though some may view these changes as threatening, it is also liberating. Museums now help people make meaning from art and objects in entirely new ways. According to this participant, the museum has become more of a hub and less of a physical resource on which people draw. The development of Web 2.0 technology has also provided museum professionals with new ways of thinking about connecting with their publics online and involving these publics in the museum environment. An interviewee explained that it is "no longer about the visitor in the life of the museum but the museum in the life of the visitor," meaning that, with the interactive Web and new levels of connectivity, museum visitors access the museum before they visit it physically, and continue to visit the museum website after they return home.

Some interviewees urged new museum professionals to rethink the very role of museums, and to consider new technological projects, not just as tools, but also as museums in themselves. Online exhibits, for example, serve as museum spaces; websites are no longer an extension of the museum, but represent the museum as a whole, particularly to those visitors who cannot physically visit the institution.

Most interviewees agreed that museums have been transformed by the proliferation of information technology in our contemporary moment. Within exhibition halls, multi-media installations have provided the most visible manifestation of technology within contemporary museum practices. This aspect has profound implications for the manner that museums communicate with members of the public, and has fundamentally changed the ways that exhibitions are undertaken. The use of cell phones and digital cameras in the exhibition space, for example, is becoming increasingly prolific. Some museums have responded to these changes by easing up on limitations placed on these devices and even encourage the public to use their cell phones in order to access interpretive guides. One interviewee also noted changes in the use of kiosks. He suggested museums used to relegate kiosks to the corner of most museum spaces, but now technology is considered more holistically at the inception of exhibition planning. However, one interviewee pointed out that exhibition teams should include members of the IT department so these teams have adequate IT expertise at all stages of the development of new exhibitions. Museum professionals need to know how to work with kiosks and build new applications for their use; as well as when to use them and when more traditional approaches will be more effective.

An interviewee working in a collection management department suggested the greatest impact of technology on the contemporary museum environment has been in the areas of dissemination and collections management. Another interviewee indicated that almost all museum departments use a collection management system, and a variety of museum professionals use the system and contribute information about objects. The collection management system is the backbone of the museum according to this interviewee. Today far more museum personnel (curators and project managers among them) are familiar with, and have access to, institutional collections because museums have begun digitizing their collections and making them available on-line since the 1990s. Today, researchers may use multiple col-

lections databases and other information technologies to "go shopping" for artifacts. Browsing the collection has been made a feasible reality. According to one interviewee, expertise in creating and working with 3D imaging is becoming increasingly important for some museum professionals. He went on to explain that 3D authoring tools, such as a 3D camera, allow museum professionals to create a three-dimensional image of specimens based on data models. Not only does this permit researchers to access collections at a distance, but it also reduces the handling of the physical specimens by providing visitors with highly flexible digital representations. Imaging technology can increase access to the collection, while helping to preserve the original object.

One interviewee emphasized that museums are about information. In his opinion, the collection remains at the core of information creation and knowledge production, and technology should be seen as a tool or interface to facilitate access to the collection for these purposes. This particular view of technology as a tool was also reinforced by another prominent interviewee who stressed the necessity for museum professionals to both remain curious and creative in their uses of technology, and to develop a solid understanding of the subject specializations of their respective museums, be it in the fine arts, material culture, history, or other. Thus museum professionals must understand and appreciate the capacity of technological tools and the value of these tools for the museum sector. An interviewee pointed out that a collection is compromised if the link between specimens and their information is broken. Therefore, collection managers must have fundamental knowledge of information management techniques and the management and use of databases. He also suggested that the management of legacy data whether still in analogue form or in antiquated collection management systems presents special challenges.

Challenges Imposed on Museums by New Technologies

The interviewees also discussed the challenges posted by new technologies and identified the three most common challenges: (1) the cost of designing, implementing, and maintaining technology; (2) a lack of in-house expertise; and (3) information management. Many interviewees noted that the collection resides at the core of any museum's work; hence, many museums have committed to building databases or adopting off-the-shelf collection management systems, and/or strengthening metadata to facilitate access to their collections amongst researchers, the public, and other museums. Data-sharing also requires that metadata meets a specific standard of interoperability so that it can be easily migrated from one IT infrastructure to the next. Cleaning data and adding to metadata are, nevertheless, tedious and time-consuming tasks. There is no uniformly adopted shared vocabulary among museums for describing collections, nor is there a consistently deployed standard form of metadata used across all collections even within a single institution. Migrating legacy data to new formats takes time and digital asset production continues at a prodigious rate. Unfortunately, museums are not able to cope with the amount of work necessary, and its expense. Without capturing information about the physical collection, however, museums will not be able to fully participate in the possibilities offered by Web 2.0 technologies such as interoperability between diverse cross-institutional data resources.

Museums are undergoing a conceptual shift that many believe has been triggered by the advent of networked computer technology. Traditionally, collections have fallen under the exclusive domain of the curator; however, as previously noted, information technology allows each collection to be shared, at least as parcels of metadata and digital images travelling across the Internet. While the mission of the museum will continue to be the preservation of its collection, opening up access has changed the focus of the museum from its preservation function to its interaction and engagement with the public. This shift in attention has a number of ramifications for museums themselves. First, the public is not always comprised of savvy technology users. From young people to the elderly, much of the general population suffers from poor computer and information technology literacy skills. For example, few people make a distinction between the deep Web and the information retrieved by a Google search. In addition, museum educators often work with primary and secondary schools, which are chronically under-funded, and, to date, often do not have access to high-speed Internet connections or software packages needed to operate Web 2.0 technologies. Museums must, therefore, strike a balance between low-tech and high-tech services to the public. Second, museums are perceived as authorities on the collections that they preserve. Interviewees point out that bad data or poor information made available on the Web reflects poorly on the subject expertise of museum professionals. At the same time, other interviewees noted that the public demands access to data that museums simply have not had the time or money to properly review or types of data that they do not have the resources to capture. One interviewee notes that most users do not ask for 100% accuracy, and if we wait until all information is 100% accurate and available museums will never upload their material

to the web. Finally, the pressure to serve the public has led many museums to pursue IT trends without understanding the costs, benefits, opportunities, and risks associated with these new technologies.
The proliferation of IT has resulted in increased rates of format obsolescence and, as a result, museums now face considerable challenges when attempting to preserve digital assets. One interviewee commented that digital photographs are sometimes mistreated by museums because they are perceived as ephemeral objects, saved en masse onto CD-ROMs and other media storage devices, which are sitting haphazardly on shelves. This is quite unlike "tangible" film negatives and analogue photographs that are being carefully preserved in the vaults. The transition from tangible to ephemeral objects is also forcing museum professionals to re-consider basic concepts of traditional museology. For example, if a digital image can be copied easily and effectively, and disseminated widely via the World Wide Web, which version of this image is the authentic record? Furthermore, how does this scenario impact the "authentic museum experience" for the public?
New media art is one area requiring particular attention, especially with respect to how the cycles of life of ephemeral objects are recorded. Organizations like the Fondation Daniel Langlois pour les arts, la science et la technologie in Montréal , whose fundamental mission is research and preservation, has partnered with many museums to further this developing area. Programs such as the Variable Media Network , undertaken by an alliance of universities and museums from 2004-2009, are the products of such work.
One interviewee stressed the importance of preserving the pertinence and authenticity of the museum experience, and this in light of a changing landscape of new technologies. Museum professionals must think carefully about their use of technology, recalling at all times the mandate of the institution and the needs of the institution's visitors. This, she observed, requires a great deal of maturity on the part of the institution. She cautioned that the content of the museum should always be considered the foundation, while the technological tools are always only the gateway
A further interviewee spoke of yet another challenge to the museum posed by the use of technology when evaluating the validity of user-added content to museum information. Museums, he argued, place a seal of quality on information that they produce, however when members of the public are invited to add images/tags/texts to this information in online environments, it becomes increasingly important to differentiate between voices, and to identify who has added what. With the growing developments of cybermuseology and exhibitions developed for the web, he argues that we must distinguish between institutional and public additions to content so as to avoid the pervasive dilemma facing the wider web: large content, but questionable validity.
Perhaps the most important challenge, however, is convincing management to understand both the benefits and limitations of technology within a dynamic working environment with conflicting priorities. Information technology is not a core function of the museum, and therefore it does not always receive adequate funding. Few museums can exist today without a level of technology and technological support. Acquiring expertise that understands both the museum context and new technologies, whether in-house or outsourced, is key to building a solid information management program and developing innovative and interesting content for the public.

Museum Professionals

As one interviewee noted, the museum profession is facing a generational shift. As older professionals retire and new professionals take their places, the museum environment is undergoing significant changes. He suggested that within the culture of museums, older generations do not tend to prioritize technology to the same extent as newer museum professionals. This he expresses this not as criticism, but merely as a reality of contemporary practices. An outcome of such practices is evident in the vision that younger generations have of the museum: their approach is broader ("transversale"), revealing a greater understanding of the many functions of the museum as a collective. This is especially true in small institutions, where professionals are encouraged to undertake many different roles.
Interviewees suggested that Generation Y tends to perceive technology as a ubiquitous part of the professional environment, not as an add-on, but rather, as inherent in day-to-day work. The challenge is for new Generation Y professionals to stress the importance of new technologies and, more specifically, social networking and other collaborative technologies, to older management. The displacement of older professionals by a younger generation is also evident to others; one interviewee mentioned that there has been a noticeable decline in the number of museum professionals seeking training in specific technologies, indicating that professionals are either entering the field already trained, or have become accustomed to self-directed learning. Several interviewees, however, warned that it is dangerous to assume that all Generation Y professionals will come to the field well versed in new technologies. While it might be true that this

cohort of professionals has been more exposed to digital technologies during their schooling both formal and informal (e.g. gaming), this should not be equated with having the ability to critically evaluate, select or use new technologies.
Another change that was noted by interviewees was the shift to project-based business and collaborative work. According to one interviewee, museum professionals are entering the field more accustomed to working in groups, drawing on the expertise of many to perform complicated tasks. As a result, museums must respond to this conceptual shift in work organization to ensure that new professionals can perform in ways with which they are familiar. Institutions, for example, might introduce Intranets to facilitate information-sharing, or chat programs to allow for geographically dispersed real-time communication. Social networking tools are becoming more commonplace within the museum workplace, as more professionals are accepting these technologies as solutions for collaborative project-based work.

Conclusions

The interviews with sixteen senior North American museum professionals underscored the vital role that information technologies have played in the transformation of their institutions over the past two decades studies in Europe [e.g. 3, 7] have produced similar conclusions. While the core mission of museums may have changed very little, the activities and operational requirements associated with expanding collections from primarily physical, to increasingly digital objects has necessitated the marshaling of customized resources and a new set of knowledge and skills. Whether collecting, curating, educating, programming, marketing, communicating, or fund-raising, all require, not only some basic engagement with associated hardware and software applications, but also a more critical understanding of the inherent strengths, opportunities, and deficiencies of information technologies to support the present and future direction of the institution. Addressing this level of required IT literacy, as well as a perceived generation gap in engagement with new technologies and tools, such as mobile devices, imaging systems, and Web 2.0 and social networking applications, demands timely, focused, ongoing training at all stages of an individual's career within museums. This is a conclusion that chimes with the results of other international studies [7]. The threshold for participation in the digital environment continues to rise and this has implications for educational needs of museum professionals. As the 2004 The Future Digital Heritage Space: An Expedition Report, made evident the diversity of mechanisms from the technologies that support the intelligent ambient landscape to those that underpin multimodal interaction such as virtual and augmented reality will increasingly change the relationship between the museum and its audiences [4]. For this to happen continuous educational learning opportunities must become a core part of the life of museums and museums professionals.

Acknowledgement

The authors would like to thank Anne-Marie Millner, Irene van.Bavel, Madeleine Lafaille and the Canadian Heritage Information Network (CHIN) for their help with this research. CHIN provided substantial funding for this research through Research Contract. No. 45247369

References

[1] Parry, Ross (2007). Recoding the Museum: Cultural Heritage and the Technologies of Change. London: Routledge.
[2] Bearman, David and Kati Geber (2008). "Transforming Cultural Heritage Institutions through New Media," in Museum Management and Curatorship, 23(4): 385-399.
[3] Geser, Guntram and Mulrenin, Andrea. (2002). The DigiCULT Report: Technological Landscapes for tomorrow's cultural economy Unlocking the value of cultural heritage. European Commission: Directorate-General for the Information Society http://www.digicult.info/pages/report2002/dc_fullreport_230602_screen.pdf
[4] Geser, Guntram and Pereira, John (2004), The Future Digital Heritage Space An Expedition Report, (Salzburg: DigiCULT, http://www.digicult.info/downloads/dc_thematic_issue7.pdf
[5] Loran, Margarida (2005). "Use of Websites to Increase Access and Develop Audiences in Museums: Experiences in British National Museums". In: CARRERAS, Cèsar (coord.). "ICT and Heritage" [online dossier]. In: Digithum. No. 7 http://www.uoc.edu/digithum/7/dt/eng/loran.pdf.
[6] Marty, Paul (2008). "Information Representation: Representing Museum Knowledge," in Museum Informatics. Paul F. Marty and Katherine Burton Jones, eds. New York: Routledge, pp. 29-34.
[7] Ross, Seamus (2001), "Budgetary Suicide at the Altar of ICT", (London, Heritage Lottery Fund,) http://www.hlf.org.uk/nr/rdonlyres/50742b80-89d2-4e6e-8a71-2469a3bcda6c/0/needs_ict.pdf
[8] For example see. Ross, Seamus, Donnelly, Martin, Dobreva, Milena, Abbott, Daisy, McHugh, Andrew and Rusbridge, Adam, (2005), Core Technologies for the Cultural and Scientific Heritage Sector, DigiCULT Technology Watch Report 3, European Commission, 296 pages. ISBN 92-894-5277-3 http://www.digicult.info/downloads/TWR3-highres.pdf. For other examples see http://www.digitcult.info

Abstract
European museums and other cultural institutions host rich collections that have ability to attract EU citizens and tourists. Cultural objects, e.g. paintings, in these collections are related in many ways and in many cases they refer to same underlying concepts, people and places. The Cultural Heritage Knowledge Exchange Platform, SMARTMUSEUM requires that these collections are interoperable over cultural and language barriers, and provides a mobile publication channel for collections.
Keywords: semantics, mobile phones, tourism, personalization

Introduction

It has been argued that museums should publish their activities, collections, services, and products in cooperation with cultural tourism agencies (Mulrenin, 2002). The Cultural Heritage Knowledge Exchange Platform, SMARTMUSEUM, is a platform for innovative services enhancing on-site personalised access to digital cultural heritage through adaptive and privacy preserving user profiling. Using knowledge bases, global digital libraries and visitors' experiential knowledge, the platform makes possible the creation of innovative multilingual services for increasing interaction between visitors and cultural heritage objects in a future smart museum environment, taking full benefit of digitized cultural information. In this paper we present components needed to realize a knowledge exchange platform for SMARTMUSEUM needs. The annotation framework enables managing of semantic annotations of museum collections. The SmartMuseum recommendation web services use relations defined by ontologies together with user profile and contextual information to provide search and recommendation for a user of SMARTMUSEUM. These web services can be used via mobile and web interfaces.

The Content Architecture of SMARTMUSEUM

The vast majority of museums hold a legacy database with cataloguing cards for owned objects. In IMSS case (Institute and Museum of the History of Science) the database contains also some ontological data. The ending result needed from the SMARTMUSEUM application is a set of RDF triples complying to an ontological schema. A clear vision of the target ontology is strictly mandatory especially in a commercial system. Since content providers can have different items and different objectives, small adjustments to the target ontology can be requested for a single implementation, but in general there should be global target ontologies available that all organizations may use.
Next, we will analyze some representative cases about how to map existing metadata to the SMARTMUSEUM ontology. Since not all the required concepts and instances were available in existing ontologies, an integration and an upgrade methodology had to be established to minimize the manual intervention of mapping. In the IMSS case the preexistent database holds information and links among information. For this reason IMSS added SMARTMUSEUM specific ontological annotations to the existing database. This implied a semi-automatic mapping and a manual addition of external references to Getty vocabularies [3]. In the second phase IMSS realized an automatic extractor that used XML-based configuration about what and how to extract from the dataset. The outcome of this process is directly usable by the SMARTMUSEUM system.

SMARTMUSEUM Recommendation Service

Profile retrieval

The first phase in the recommendation is to retrieve a profile that matches the user's current context. Retrieving is done by mapping the user's location, determined by GPS to ontological concepts. This is done by expanding the query of the single coordinate point to cover a circular area within some radius r, say, 1000 meters. This is done in two steps. First,

Chiara Cirinnà and Maurizio Lunghi (edited by), *Cultural Heritage on line: Empowering users: an active role for user communities. Florence, 15th - 16th December 2009* ISBN 978-88-6453-184-7 (print) ISBN 978-88-6453-187-8 (online)

a simple bounding box is created where each of the edges of the box have a distance r from the user's location, i.e. the distance to the "sides" of the bounding box from the given location is r. In the second phase, places outside the radius r are further pruned away from the results. This results into an ontological resource representing the position information. We use the likelihood of a context generating a certain triple. It can be observed from the relative frequencies of the profile entries. For example, if a user profile contains tags for triples about Italian paintings in the context of Helsinki, say 10 times, and triples in Helsinki in total 20 times, the P (<sm:Painting, sm:manufacturedIn, place:Italy> | <rdf:Resource, sm:userLocation, place:Helsinki>) would be 10/20 = 0.5. Because we have the negative or positive votes for the triple we calculate the average of the votes of the triple in the given context and multiply it with the probability of the triple in the context. The contexts in which the observations are done can be very sparse. Therefore, we use Laplace (i.e. add one) smoothing [4] to shave a share of the probability mass to contexts for which no observations are available. In this way, we can observe some probability for every triple even if it has not been tagged in the specified context.

Recommendation retrieval

Recommendation retrieval is performed by using the query constructed from user profile and context [2]. Based on the earlier phases we have a set of profile triples that each has weight. Each triple may be expanded using query expansion to multiple triples, that each has the weight of the original triples. This is done by including all triple combinations having a Wu-Palmer value higher than a fixed constant. We have set this value to 0.85 by error and trial. As a result we have a set of triples each having a weight.

We can now define the retrieval as a two step matching procedure that utilizes the spatial constraints and a scoring function used to calculate the cosine similarity [6] in vector space model [5], where vectors are formed by using a triples times documents matrix. Further, we cluster the best 300 objects using independent component analysis (ICA). This makes it possible to reveal different viewpoints to the data and avoid over-specialization. For example, if the user has a very strong interest in Italian paintings and a light interest to scientific instruments and telescopes, a traditional retrieval system would only rank Italian paintings high. By clustering, we are able to build three interest clusters, one for Italian paintings, one for telescopes and one for other scientific instruments. Finally, items from each cluster can be included into the final recommendation list.

Mobile access to SMARTMUSEUM

Mobile access to SMARTMUSEUM is based on two scenarios: the inside scenario - user visits a museum and the outside scenario - user walks around the city, looking for (outdoor) points of interest (POI). In both scenarios the user is equipped with a PDA or a smartphone as a main device for user's positioning and for presenting recommendations and multimedia (A/V, Text To Speech) information. SMARTMUSEUM mobile access is based on five major features:

- User is requested to enter the expected visit duration and a purpose of visit from a predefined list.
- User can acquire contents describing recommended objects by clicking on URLs displayed by the mobile device. According to user's profile preferences, multimedia files or text-to-speech contents are automatically launched when they are available.
- Each object has a unique URI, a URL is usually stored into a RFID tag for indoor scenario, or GPS coordinates are used for outdoor POIs. When looking at an interesting indoor object, user scans the RFID tag attached to the object to get information on the mobile device. Each user browsing action is logged.
- User has an opportunity to rank each page (physical objects and content pages).
- After the end of visit ranking and log information is automatically sent into the SMARTMUSEUM profile server.

Mobile user positioning and interest monitoring

Mobile user monitoring is performed (1) for positioning and (2) for discovering user's interest in order to process statistical recommendation. Outdoor user positioning in SMARTMUSEUM context is used for determining location for nearby POI search and for determining location context for semantic recommendations. The SMARTMUSEUM solution supports GPS (WGS84 coordinates) and mobile network cell based outdoor localization. For indoor scenario the objects are equipped with 13,56MHz (ISO144443A) RFID tags used for content triggering and user positioning as well.

RFID reading

A unique solution was developed for RFID tag access. Existing solutions are mainly based on unique hardware ID of the tag. Using this ID for referencing an object would require changes in the SMARTMUSEUM database when tags are replaced. For this reason we made tag data area usable for content. Two solutions have been studied and evaluated: 1) store html contents directly on tags, 2) store an object URI on tag and to retrieve information through wireless network. We made several measurements with 4K tags. Our experiments showed that retrieval of a small html page (3071 octets, without image and CSS style) requires 5 sec. The reading of a simple URI takes only 150 msec. Here the problem is that the user must keep the device near the tag until the end of reading. With the first approach, there is a risk of retrieving partial information. In the second solution, tag reading is almost immediate so that user behaviour is not constrained. Information are automatically loaded and displayed after tag reading. We made some performance measurements with the server of IMSS. It takes about 1.5 sec to read tag and to load a simple html page (without image of object). Finally the second solution has been fully integrated in the current release of our software.

Monitoring user interest feedback

Receiving pertinent user feedback is a crucial factor to process further statistical recommendations for both objects and content pages. Browser activity logging is a widely used method for web content relevance evaluation, especially from server side. In many studies page access duration is taken into account. However, time counting is quite unsuitable for mobile users, because the content access is highly fragmented. So it is difficult to detect idle periods when user, for example, is moving from one object to another one. The second issue is that since content is fully distributed, the activity monitoring cannot be performed in a centralized manner. But in SMARTMUSEUM context, most of the available information is split into separate html pages. In this way it becomes possible to infer user's interest for each museum object and atomic content piece. Mobile software stores all visited URL, and/or visited object URI. And a score is associated to each URIs and URLs.

After the analysis of user scenarios, a combined scale of manual preference input and implicit monitoring of preference and behaviour have been proposed and implemented (see Figure 1):

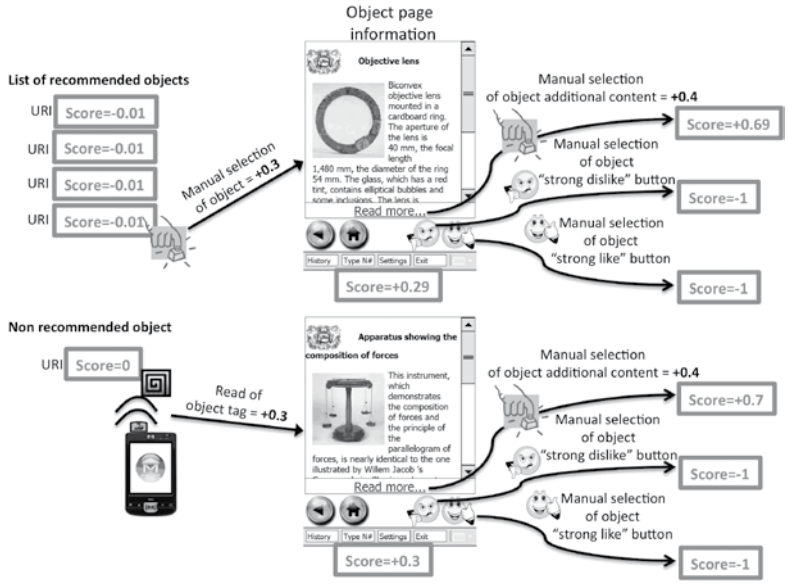


Figure 1: Rating of a SmartMuseum object

- score = -0.01: user receives recommended object. If the user shows no interest, and leaves associated recommended link unused, this initial score remains unchanged.
- score = score + 0.3: user fetches basic information about the recommended object or user reads object tag.
- score = score + 0.4: user requests more information about an object to receive a list of additional content URLs.
- score = 1: strong like, manual input on user interface.
- score = -1: strong dislike, manual input on user interface.

Currently the user device client software is implemented for Windows Mobile (WiMo) and Symbian operating system platforms. Two screenshots of WiMo user interface optimized for larger touchscreens are presented in Figure 2 (recommendation list and object information page). Our objective is to minimize user interventions. Upper part of screen is an embedded Web browser window, on bottom part rating buttons can be used. On this screen area multimedia and text-to-speech controls appear automatically when content html page includes respective hidden tags.

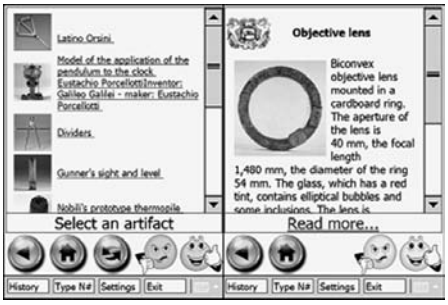


Figure 2: Mobile device main UI (WiMo)

Conclusions

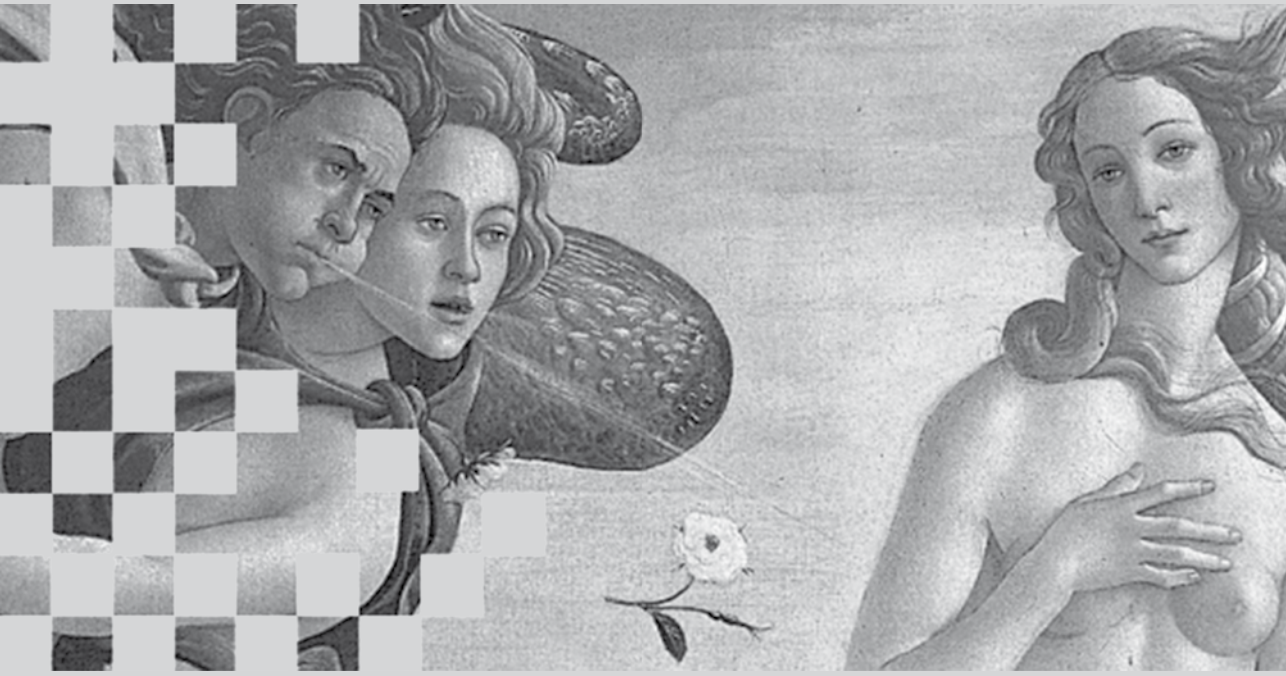
SMARTMUSEUM is a versatile knowledge exchange platform for hosting and publishing museum collections and POIs for user's of mobile phones and PDAs. In this paper we described different components used for realizing the SMARTMUSEUM.

Acknowledgements

The research described in this chapter was done in the EU project SMARTMUSEUM supported within the IST priority of the Seventh Framework Programme for Research and Technological Development. Partners of the SMARTMUSEUM include Apprise (Estonia), National Institute for Research in Computer Science and Control INRIA (France), Helsinki University of Technology (Finland), Royal Institute of Technology in Stockholm (Sweden), Webgate JSC (Bulgaria), Heritage Malta (Malta), Institute and Museum of the History of Science (Italy) and Competence Centre of Electronics-, Info and Communication Technologies ELIKO (Estonia).

References

[1] Mulrenin, A. (Ed.). (2002). The digicult report. technological landscapes for tomorrow's cultural economy. unlocking the value of cultural heritage. executive summary. Luxembourg: European Commission.
[2] Tuukka Ruotsalo, Eetu Mäkelä, Tomi Kauppinen, Eero Hyvönen, Krister Haav, Ville Rantala, Matias Frosterus, Nima Dokoohaki and Mihhail Matskin: Smartmuseum: Personalized Context-aware Access to Digital Cultural Heritage. Proceedings of the International Conferences on Digital Libraries and the Semantic Web 2009 (ICSD2009), September, 2009. Trento, Italy.
[3] http://www.getty.edu/research/conducting_research/vocabularies/
[4] Christopher D. Manning and Hinrich Schuetze. Foundations of Statistical Natural Language Processing. The MIT Press, 1 edition, June 1999.
[5] G. Salton, A. Wong, and C. S. Yang. A vector space model for automatic indexing. Commun. ACM, 18(11):613–620, 1975.
[6] Gerard Salton and Christopher Buckley. Term-weighting approaches in automatic text retrieval. Inf. Process. Manage., 24(5):513–523, 1988.



Parallel sessions II

Sustainable policies for digital culture preservation

* This introduction is partially based on the keynote contribution presented at the conference *Perspectives on Metadata,* held in Vienna, 12-13 November 2009,
https://fedora.phaidra.univie.ac.at/fedora/get/o:45908/bdef:Asset/view.

A PREMISE

This introduction will be dedicated to present a common perspective on digital preservation by assuming that basic requirements for its success have conceptual and organizational nature, as increasingly recognized by the literature and the research outputs in the field. The metadata for preservation, the early adoption of adequate formats, controlled methods and good technical standards for acquiring digital resources play their role for ensuring the sustainability of the function, but they need to be included within a comprehensive and convincing intellectual framework and well state responsibilities. If the specific applications and related tests are not included within a systematic and robust theoretical infrastructure, the fragmentation is not avoidable and the risks for failure increase. This is why we have to put the accent on the relevance of the main goal and principles of the entire system (the defense of its trustworthiness and credibility) and its roots (the conceptual framework) and on the correct identification of responsibilities and procedural rules (the custodial environment as a chain of custody and its certification), both required for developing new products and implementing the existing solutions. This introduction will start from two assumptions:

- first of all, the challenges still open, specifically for handling the creation and preservation of digital resources depends on the recognition of their dynamic nature and the related need for handling as part of continuing and ongoing processes: the digital world offers a rich series of tools for the identification and capture of metadata and information on the basis of their position and encoding: they can appear as attributes of the resource itself, i.e. in the face of the digital object, as logical and physical components of its form, they can play as external elements (i.e. in a database system), but they also can act as implicit information within the procedural, technological or juridical contexts and they have to be captured and, even more, understood and maintained;
- a pragmatic effort is required but it must be strongly rooted on consistent theory and principles specifically if we want to play with advanced technologies): it must be able to combine the best models for interdisciplinary approach, to avoid a useless overloading of detailed but not always useful information and to take into account in the application the promising outputs of the most recent research projects (PLANETS, CASPAR, INTERPARES, PREMIS just to mention those already known for their successful achievements and presented and discussed in this conference.

InterPARES is here considered as a conceptual framework thanks to principles, policies and procedures tested in many case studies and based on a consistent dictionary. The OAIS standard is recognized as a reference model for information architecture but also – specifically in the CASPAR project - as an implementation system. The guiding principle of CASPAR has been the application of the OAIS Reference Model to research, develop and integrate advanced components to be used in a wide range of preservation activities and to create a specific framework as a software platform for preservation that enables the building of services and applications that can be adapted to multiple areas, specifically to cultural, scientific and performing arts domains (that is dynamic sectors which require very complex and really evolving solutions).
CASPAR and PLANETS conceptual models have included multiple relevant results achieved in the field of preservation in the course of the last decade research efforts: the principles of InterPARES itself, the OAIS general framework, the checklist for auditing digital repositories developed in the TRAC report (Trusted Repository Audit Checklist) and in the RAC recommendations (Repository Audit and Certification), the PREMIS schema developed as metadata for digital preservation, the ISO standard CIDOC (Conceptual Reference Model) for developing ontologies and mapping metadata schemas with semantic functionality . The motivation was the creation of digital repositories and the develop-

Chiara Cirinnà and Maurizio Lunghi (edited by), *Cultural Heritage on line: Empowering users: an active role for user communities. Florence, 15th - 16th December 2009* ISBN 978-88-6453-184-7 (print) ISBN 978-88-6453-187-8 (online)

ment of framework and services for preservation based on an integrated approach to be applied to differentiated and complex archival and information systems.
The contributions presented in this session have made constant reference to these results in the specific effort for developing concrete domain-centric solutions.
The definition (and the agreement on the role) of a conceptual framework for ensuring both the consistency and the efficiency of the digital repositories requirements and of the preservation action in terms of policy, procedures and responsibilities is a key basic issue, a condition to transform into an interrelated approach the individual solutions based on metadata identification and extraction or on the development of persistent identifiers criteria as it will be illustrated and discussed further in the course of this session.
The solidity of this analysis and chiefly the consistency of its implementation need some general statements. Specifically we could/would agree at least on the fact that the handling of digital assets as reliable, accurate and authentic heritage implies the clarification of the principle of trustworthiness.
If we look at the applications developed at national level, in most cases we could see continuing and exacting attempt for integration of principles and tools as outcome of research projects and standards development. But the fragmentation is difficult to overpass and it is even more complex to build a organic scenario.

The conceptual framework and the principle of trustworthiness for digital preservation

The information and record preservation is increasingly based on concept of trust, specifically if the environment becomes digital.
First of all, it is suitable to share the definition of this term and clarify the connection between the concept of trust and the nature and quality of the digital heritage to be preserved, because the questions related to the metadata collection but also those concerning the responsibilities and the technological and organizational contexts for preservation are involved in this analysis and cannot be used conveniently and efficiently without this clarification.
In the dictionary (Merriam-Webster, s.v.) trust is identified as "a charge or duty imposed in faith or confidence or as a condition of some relationship", a sort of "glue which binds that relationship together"[1], whose ingredients have to be identified and described for effectiveness of the custody.
The custody can play successfully its role if all the elements and activities involved in this function can imply or presume a trustful handling and accomplishment.
According to the recent CCSDS guidelines, still published as draft (Recommended practice: Requirements for bodies providing audit and certification of trusted digital repositories, http://wiki.digitalrepositoryauditandcertification.org/bin/view/Main/ReqtsForAuditors) the trust is at the basis of the certification process and at the centre of the whole process for providing solidity and efficiency to the curation action in the digital world. It involves a large community:
"to give confidence to all parties that a management system fulfils specified requirements. The value of certification is the degree of public confidence and trust that is established by an impartial and competent assessment by a third-party. Parties that have an interest in certification include, but are not limited to
the clients of the certification bodies,

- the customers of the organizations whose management systems are certified,
- governmental authorities,
- non-governmental organizations, and
- consumers and other members of the public".

It requires the identification of reference principles able to inspire confidence. This kind of principles includes (according to the CCSDS report):

- "impartiality,
- competence,
- responsibility,

[1] 1 See Jennifer Borland, Trusting Archivists, in "Archivi & Computer", 2009, 1, pp. 95-106.

- openness,
- confidentiality, and
- responsiveness to complaints".

Each single attribute should be evaluated and transformed into procedures, rules, tools and metadata collection in a way to provide frames and contents for the evaluation of requirements and the recognition of the quality of digital repositories and their management and preservation systems.
Specifically, a more detailed exam of the core definitions could be of help for investigating the efficient use of metadata finalized to

- foster the credibility of the repository as trustworthy custodian on the basis of its capacity of securing integrity and authenticity of their digital contents through a standardized accumulation of descriptive and management information,
- control the cost of descriptive function "by using a simple [and standardized] encoding scheme and by ingesting metadata on transfer from public sector institutions",
- enlarge the range of interrelations by "exchanging finding aid metadata with metadata harvesters from all kinds of communities".

We do not have here time for this analysis, but it is important to recognize, within this perspective, the risk of fragmentation in the collection of all these information elements[2] and the low capacity of the present schemas and standards to document comparatively processes and describe them with an holistic and dynamic approach, the only one capable of dealing with the continuing evolution of the technological complexity. Of course, this last aspect, the most crucial for preserving the digital resources, requires the design of the digital preservation work as a chain of custody based not only on content identification, description and protection but also and with an increasing emphasis on the requirements for certifying institutional dedicated repositories, common policies and well defined and documented responsibilities.

THE CHAIN OF CUSTODY: REQUIREMENTS, POLICY, RESPONSIBILITIES

"The enduring trustworthiness of our documentary heritage is becoming a central responsibility of its designated custodian"[3], as neutral third party on the basis that "it has no reason to alter the records and no interest in allowing others to do so, and must have the knowledge necessary to implement procedures that ensure the integrity and accuracy of the records"[4]. This assumption is today at the centre of a common effort made by the professionals involved in digital documents and in digital forensics, all of them persuaded that the core concepts concern the creation of a multilayer approach able to verify the integrity and authenticity of the resources at various levels of analysis:
on the basis of the elements on the face/in the form of the resource and its attributes and metadata,

- from the circumstances of its maintenance and preservation: "an unbroken chain of responsible and legitimate custody is considered an insurance of integrity until proof to the contrary"[5],
- from the integrity of essential metadata related to the resources handling and preservation as a further requirement for attestation of integrity and authenticity (individuals/offices involved, indication of annotations, of technical changes, of presence or removal [and the related time] of digital signature and other digital seals, the time of transfer to a trusted custodian, the time of planned deletion, the existence and location of duplicates outside the system,

[2] See Kai Naumann, Christian Keitel, Rolf Lang , "One for Many: A Metadata Concept for Mixed Digital Content at a State Archive", The International Journal of Digital Curation, 2009, 2, http://www.ijdc.net/index.php/ijdc/article/viewFile/120/123: "It is the diversity of these objects which represents the key challenge in devising a metadata concept to describe, preserve and distribute them. They all need to be located on the existing finding aid system, regardless of their media format". See also Pikka Heutonnen, "Creating Recordkeeping Metadata", Atlanti, 19 (2009), pp. 67-76.

[3] L. Duranti, From Digital Diplomatics to Digital Records Forensics, in print.

[4] Ibidem, with specific reference to Bernard D. Reams Jr., L. J. Kutten, and Allen E. Strehler, Electronic Contracting Law: EDI and Business Transactions, 1996-97 Edition (New York: Clark, Boardman, Callaghan, 1997), p. 37.

[5] L. Duranti, From Digital Diplomatics to Digital Records Forensics, cit.

- as inference on the basis of the trustworthiness of the record/document/information system in which the records/documents/information exist.

As Luciana Duranti has recently clearly expressed, "the authenticity…is a removable responsibility, as it shifts from the creator's trusted …keeper, who needs to guarantee it for as long as the record is in its custody, to the trusted custodian, who guarantees it for as long as the record exists"[6] .
If the framework and some basic principles seem today accepted and constitute the basis for the future implementation, some relevant details stay undetermined.

What is still missing

1. consistent and accepted terminology and definitions used across domains and requested to be well understood beyond the professional communities involved in digital curation environment with specific reference to the fact that:
- definitions related to the attributes of preservation are not clearly expressed and present dangerous ambiguities[7],
- new terms or the revision of traditional expressions (i.e. significant properties[8]) can produce dangerous misunderstanding;
- OAIS glossary has still inconsistencies even if the standard is a fruitful framework for implementing digital curation/preservation environment and has the ambition and the capacity to define concepts for a general frame: the new version (under final approval) has not been able to solve all the uncertainties even if a serious improvement is easily recognizable.

2. development of interrelations and concrete and open cooperation among relevant projects and standardization process (like PREMIS, InterPARES, PLANETS, CASPAR, DRAMBORA, RAC, CIDOC) with the aim of building an interoperable framework and diminishing the present fragmentation for a better orientation of the users.
As a consequence:
3. integration of models, schemas and business solutions to be developed in the application scenarios for handling relevant tasks as:
- authenticity and its presumption,
- storage systems in independent environment,
- automated metadata extraction: on this last point, some efforts have been made recently, but the results are slow and not enough convincing. The time is not enough to enter into details. Two recent contributions to the field could be taken into account: Kim-Ross research on automated genre classification and the FinnONTO project developed in Finland[9].

The complexity and the contradictions of the digital world could have two opposite consequences, as directly experienced by many e-government legal frameworks and preservation projects: frustration and inactivity on one side, free attitude for creating, testing and supporting innovation on the other side without avoiding or hiding difficulties. Of course the last possibility requires capacity, courage and most of all confidence on the professional accumulated knowledge. The session has offered the opportunity to share ideas and increase the quantity and the quality of this knowledge in one of the most complex and relevant task we have to face, rich of promises and contradictions. One more reason to thank the organizers for this event and all of contributors for their efforts.

[6] Ibidem.
[7] M. Day, Preservation metadata, http://www.slideshare.net/michaelday/preservation-metadata.
[8] The definition of significant properties is emblematic of the pointlessness of this new term: "the characteristics of digital objects which must be preserved over time in order to ensure the continued accessibility, usability, and meaning of the objects, and their capacity to be accepted as evidence of what the purport to record" (see Andrew Wilson, but also InSPECT - Investigating the Significant Properties of Electronic Content over Time). The term seems to concentrate what common sense normally does.
[9] Y. Kim, S. Ross, "The Naming of cats. Automated Genre Classification", International Journal of Digital Curation, 2 (2007), 1, http://www.ijdc.net; Pikka Heutonnen, "Creating Recordkeeping Metadata", Atlanti, 9 (2009), pp. 67-76. For the FinnONTO project see www.seco.tkk.fi.

Abstract
This paper proposes a reflection on the points of convergence between the fields of digital preservation and digital accessibility, in terms of both research and development. The two areas have little exchange between them. But, if we look more closely, we find numerous elements – such as objectives, procedures and unresolved problems – that coincide.
Starting with objectives, each area strives to serve users who, at first glance, are quite different: digital preservation is aimed at future users that will use digital platforms that are still unknown, whereas accessibility focuses on current users with disabilities or within disabling contexts. But on closer look, there are parallels between the two groups of users. In both cases there is a considerable lack of understanding about the true needs of users and many unknowns about their technical usage requirements.
As to procedures, the standards for preservation (ISO 14721:2002 – OAIS) and accessibility (CWA 15778:2008 – Document Processing for Accessibility) share obvious similarities. Both propose a model in which there are entry formats, internal formats, and output or dissemination formats. The criteria for format selection in both fields are frequently quite similar.
Finally, there are common, unresolved problems. In the field of preservation a debate has long existed about which "significant properties" need to be preserved, whereas with accessibility, in the absence to date of serious consideration about elements such as emotional aspects, this debate is just beginning.
In conclusion, there is an evident need and rationale for establishing bridges between the two fields in order for them to learn from one another. If they join forces, it is quite possible that common solutions can be found.
Keywords: Digital preservation; Digital accessibility; Long term access.

Introduction

The goal of digital preservation is to allow documents produced in the past to be accessed in the future. For this reason "access" – or "accessibility"– of preserved documents is one of the recurring themes in this area. At the same time the term "digital accessibility" has another meaning: it is the combination of techniques that make it possible for digital documents to be used by anyone, regardless of possible disabilities: vision impairment, motor difficulty, deafness, etc. The aim of this paper is to relate the "accessibility" concept of digital preservation with this second meaning and comment on similarities and differences between the two areas.
Digital preservation and accessibility are two distinct areas, but we believe that it is worth noting the important similarities in the problems that each seeks to resolve. As such, we believe that the way in which solutions are sought in one area can, at the very least, shed light on issues under consideration by the other. For example, both areas have addressed how to select and maintain important document features for subsequent access: to future generations, in one case, and to users with sensory disabilities, in the other. Another relation between the two areas is the uncertainty surrounding the needs of real users, either because they are future users and we do not know what technology they will be using; or because the technology for assisting them currently advances at such a fast pace that their needs adjust continually to the ever-increasing capacity of new systems.
In spite of these points in common and the fact that both areas work with the same elements – digital objects - they do so in separate ways, in terms of the persons, institutions and standards that are devoted to them.
One example that reveals how preservation and accessibility are not marching in unison is that of open repositories of scholarly material, promoted by universities and other research institutions: given the importance of the stored content, preservation aspects are being given attention but, paradoxically, little attention is being paid to the current accessibility of these same documents. [1]

Chiara Cirinnà and Maurizio Lunghi (edited by), *Cultural Heritage on line: Empowering users: an active role for user communities. Florence, 15th - 16th December 2009* ISBN 978-88-6453-184-7 (print) ISBN 978-88-6453-187-8 (online)

BENEFICIARIES

On a conceptual level both preservation and accessibility share the broad goals of working to serve all types of users, but the reality does not reflect this ideal. For example, the directives for the accessibility of web content recognize explicitly that they do not include users with cognitive disabilities. [2] Also, even if documents are created following the existing standards, their accessibility is not guaranteed. Some producers in targeting a specific audience may decide that a given property is not essential, even though this will cause the product to be inaccessible for other groups for which the eliminated property may be very important. [3]

Even though generic techniques exist that permit specific documents –or parts of documents—to be accessible to all users, in many cases it is absolutely necessary to know the potential audience in order to apply the most relevant solutions. For example, designing a product for the prelingually deaf may result in a sharp reduction of textual language, even though the resulting product is then ill-suited for persons with visual impairments. Similarly, in the field of commercial publishing it has proven impossible to create via a single production line a digital work that responds adequately to all situations, publishing channels, and needs. Thus, the CWA recommendation gives examples of good practices with diverse scenarios, but it makes it evident that each situation will require solutions adapted to its own users. So in the end technical efficiency and economic viability are the parameters that determine the adoption of particular accessibility solutions.

In preservation the vision is similar, but expressed using a different terminology. As the OAIS standard itself states, the purpose of preserving digital information is to "make it available for a Designated Community" [4]: the intended future users of the preserved digital objects. And why is this so? For a very simple reason: the awareness of the difficulty –if not to say the impossibility– of fully preserving all original properties of the digital objects. Again, technical constraints and the need for economic viability lead to solutions in which only some significant properties, or essential elements, of the objects are preserved. And this poses a question –which elements are essential?– that can only be answered from the perspective of a given designated community. [5][6] Even for the experts this is not an easy matter because the choice of significant properties is subjective, making it difficult to arrive easily at agreements.

There are two main streams of thought regarding accessibility [7]: the user-centred design, more inclined to create specific solutions for different communities (the elderly, those with motor disabilities, etc.); and the universal design that promotes the idea of a single design to serve all publics. Nonetheless, both visions share the belief that documents produced with accessibility in mind will end up being better for everyone. On the other hand, with preservation there is the growing tendency to design systems adapted to a specific community of users, in which the preservation of given significant properties are prioritized over others. As a result in the future we may find documents that are valid for one community, but perhaps totally unintelligible or unusable for others.

These choices – be they related to accessibility's audience or to preservation's designated user community - lead to a renunciation of the digital object's universal applicability and can prove difficult for different sectors' experts to accept. For example, questions arise such as: Why renounce the subtitles of certain videos? Why not preserve the original typography and colour of a catalogue of artworks?, etc.

PROBLEMS

Making digital documents and computer applications accessible, as well as preserving all types of digital objects, are stimulating missions, but at the same time difficult to accomplish fully. The basic principles can clash with formidable technical, economic, and management difficulties.

The first difficulty is the broad reach of the missions: at present it is impossible to make all content accessible and to preserve all that needs to be preserved. Priorities must be established and the criteria can vary: the easiest, the most economical, the most scalable, the most heavily used, etc. Therefore, prioritization requires policies to be applied. And the other side of the prioritization coin is the renunciation: of what (for the moment, perhaps) will not be accessible or will not be preserved. The policies of prioritization are painful because implicitly they go against the global aims: some disabled persons will not have access to content that they perhaps will need; others, in the future, will not have access to specific data or testimony from our time.

Traditional accessibility solutions, such as screen magnifiers or screen readers, are built upon the applications and thus are not well integrated into operating systems and other programs. In the long run, the solution will lie in incorporating

accessibility into all phases of the development of hardware, software and content, as well as having it present in the workflow of document management. In digital preservation the use of proprietary file formats multiplies the challenges of managing their preservation, as does the plethora of existing formats. A similar reflection can be made regarding the limited support that standardised metadata schema receive from many software applications, not to mention file formats that do not admit metadata.

Legal barriers are also common to both accessibility and preservation. In the analogical world, in many countries the law protects the rights of disabled persons by setting limits to the intellectual property rights in order to allow for the publication of books in Braille. In the digital world, there tend to be fewer exceptions, a situation that leads to increased expense for the rights to publish accessible works. This has opened new fronts for the struggle to broaden rights concerning digital documents. [8] In other cases the problem does not stem from the document itself, but rather from the existence of proprietary reader software that impedes the introduction of elements that that contribute to accessibility.

The problem is similar with preservation. Laws protect the rights of copyright holders by prohibiting the reengineering or decompiling of software, or the modification of content formats, to cite three of the major techniques applied in many preservation scenarios. Certainly, in recent years there have been numerous initiatives to permit such activities in the context of preservation. But at present, many preservation-related actions currently take place in an environment of questionable legality, at the least. Similarly licenses and usage restrictions –sometimes in the form of Digital Rights Management (DRM) – are also barriers for producing documents that are accessible to all. [9] They also act as barriers for full preservation.

The timing of implementation is also important. The law in many countries protects the publication of accessible versions of textbooks for disabled students, especially the visually impaired. Nevertheless, the procedure for exercising this right can be slow, and may not conclude until after the publication of the commercial work. Therefore, users dependent on the accessible version receive the work much later than others. [10] Similarly, preservation is still seen in many scenarios as an activity taking place at the end of the "normal" life of a digital object and hence is not considered until it is time to "store" the object. At that stage, immediate treatment –for example, migration or other procedures – may be necessary, which might have been spared had the digital objects been created in accordance with preservation requirements.

Technological foundation

A comparison of the principal technologies of accessibility and preservation leads us to conclude that there are important similarities both in the procedures and recommendations promoted in each area. The major ones are: the transformation of file formats as a basic technique for facilitating present or future access; the standardization and use of structured and open formats; and the requirement to make full use of metadata.

For accessibility, the use of standards in software and open formats facilitates interoperability and, therefore, the integration of technical aids for reading the documents. The use of structured formats from the XML family facilitates the transformation of documents and, therefore, it too aids in the generation of versions adapted to the needs of different user communities. [11] DAISY is perhaps the format that is currently experiencing the greatest development along these lines, within the publishing sector. [12]

In the field of preservation, many experts have prioritized the reduction of formats used and the appropriate choice of formats for subsequent preservation. The choice of formats is frequently made during the creation, or even during use, of the document and thus remains beyond the scope of preservation actions. However, some programs encourage the use of open, interoperable and standard formats. [13] An appropriate characterization of files and the proper structuring of contents in them can also contribute towards subsequent preservation tasks, such as migration.

Whether from the vantage point of accessibility or of preservation, the volume of digital production is so great that it is virtually impossible for all files to be handled appropriately after the fact, e.g., after their creation. This leads to the recommendation that files should be standardised at their point of origin, as a means of reducing variability. It should not surprise us, then, that accessibility experts are promoting the adoption among publishers of standards and common formats, for both webs and textbooks. [14] This would enable the publishing chain to generate specific products with varying presentations and formats geared to the needs of each user. In preservation, there are many more sources of content generation, since digital objects created within the publishing world account for only a small fraction of the total

number of items to be preserved. Some current attempts for influencing how digital objects are created are centred within the public administration and some scientific fields and it remains to be seen if and how it will spread to other areas in the future.

CONCLUSION

Accessibility and preservation serve different objectives even though they act on the same types of materials. In this work we have seen some of their similarities: in strategies, in approaches to challenges, and in technological underpinnings. We have also seen that some proposed solutions are stymied by pre-existing legal conditions. Likewise we have shown how practical concerns – such as technical expediency and economic viability - can lead to actions that are frequently more limited than the respective movements' overriding aims.

These common elements lead us to believe that a greater degree of understanding between the two communities would be beneficial to both sides. Surely each could learn something from the other and, in so doing, shed more light on its own approaches. Also, collaboration would be beneficial in order to reach common objectives, such as the promotion of open and structured formats.

Finally, it is worth remembering that the two communities maintain close relations with certain stakeholders: universities, public administration and libraries. Equally important to both is the expansion of e-government as the main transforming engine for practices related to the creation and management of digital content. This common ground could facilitate points of encounter and contribute to working together towards common solutions.

REFERENCES

[1] Kelly, Brian (2006): "Accessibility and Institutional Repositories". UK Web Focus. http://ukwebfocus.wordpress.com/2006/12/12/accessibility-and-institutional-repositories/

[2] Web Content Accessibility Guidelines (WCAG) 2.0. W3C Recommendation 11 December 2008 (2008). http://www.w3.org/TR/WCAG20/

[3] CWA (2008): CEN/ISSS CWA 15778:2008 – Document processing for accessibility. ftp://ftp.cenorm.be/PUBLIC/CWAs/DPA/CWA15778-2008-Feb.pdf

[4] Consultative Committee for Space Data Systems (2002): Reference model for an open archival information system (OAIS). Blue Book. http://public.ccsds.org/publications/archive/650x0b1.pdf

[5] Cedars Project (2001): The Cedars Project Report. April 1998 – March 2001. http://www.leeds.ac.uk/cedars/pubconf/papers/projectReports/CedarsProjectReportToMar01.pdf

[6] Hedstrom, Margaret; Lee, Christopher A. (2002): "Significant properties of digital objects: definitions, applications, implications". Proceedings of the DLM-Forum 2002. Luxembourg, Office for Official Publications of the European Communities. p. 218-223.

[7] Seale, Jane K. (2006): E-Learning and Disability in Higher Education: Accessibility Research and Practice. Oxford, Routledge.

[8] Commission of the European Communities (2008): Green paper: copyright in the knowledge economy. Brussels, Commission of the European Communities. http://ec.europa.eu/internal_market/copyright/docs/copyright-infso/greenpaper_en.pdf

[9] Kramer, Elsa F. (2007): "Digital Rights Management: Pitfalls and Possibilities for People with Disabilities". Journal of Electronic Publishing, 10 (1). http://hdl.handle.net/2027/spo.3336451.0010.106

[10] Keil, Sue; Parris, Delith, Cobb, Rory; Edwards, Angela, & McAllister, Richard (2006). Too little, too late - provision of school textbooks for blind and partially sighted pupils. London, Royal National Institute of the Blind.

[11] Paepen, Bert; Engelen, Jan (2002): "Using XML as a Reading Enabler for Visually Impaired Persons". 8th International Conference, ICCHP 2002. Lecture Notes in Computer Science, 2398, p. 382-389.

[12] Kahlisch, Thomas (2008): "DAISY: an opportunity to improve access to information for all". Information Services and Use, 28 (2), p. 151-158.

[13] Arms, Caroline; Fleischhauer, Carl (2005): "Digital formats: Factors for sustainability, functionality and quality". Proceedings Society for Imaging Science and Technology (IS&T) Archiving 2005. Washington DC. p. 222-227.

[14] Lockyer, Suzanne; Creaser, Claire; Davies, J. Eric (2005). "Availability of accessible publications: designing a methodology to provide reliable estimates for the Right to Read Alliance". Health Information and Libraries Journal, 22 (4), p. 243-252.

Abstract
This paper presents the Planets Testbed, a web-based application that provides its users with a controlled collaborative environment for scientific experimentation in digital preservation. The paper gives an overview about the core concepts of the Planets Testbed and describes how the application supports the user community in preserving the digital cultural heritage.
Keywords: Planets project, Testbed, digital preservation, long term preservation

Introduction

The Planets Testbed is one of the core results of the FP6 Planets Project (http://www.planets-project.eu) which aims to create a software suite capable of addressing the digital preservation challenges that libraries, archives and the digital preservation community are currently facing.
The Planets Testbed is more than a software package – it is a central environment (consisting of software, hardware and data) for testing the performance and capabilities of tools for digital preservation. The tools are offered as web services which can be combined in complex workflows. Measurement processes are highly automated, allowing large amounts of tool evaluation results to be collected via mass experimentation.
The Planets Testbed is essentially community software dedicated to people dealing with long term preservation issues on a day-to-day basis. In the following, we will provide an overview of the Planets Testbed and discuss its role for the dedicated user community and for the preservation of the digital cultural heritage.

The Planets Testbed

The Planets Testbed Environment

The Planets Testbed provides a web-based software allowing to explore and test preservation services. This software relies on a Planets-wide, interoperable infrastructure, through which different tools can be invoked in a uniform way: the Planets Interoperability Framework. It defines the generic interfaces enabling the seamless integration of a large number of tools each of which provides a specific functionality required for performing long term preservation tasks.

The Experiment Process

Different kinds of experiments are divided into different `Experiment Types' (see section 1.3). Each experiment type of is based on a workflow which itself consists of a sequence of preservation service operations.
Using the Planets Testbed web application, the user is guided through six steps of an experiment process, as shown on the left-hand side of Figure 1. The following walk-through will use an example which might play a role in a real institutional process: The automated characterisation and migration of digital content. To be more concrete, this example refers to the migration of a single TIF file to a single JPG file, and subsequently the comparison of the properties of the input and output files.

Define Basic Properties

In the first step of an experiment, basic experiment metadata is recorded. A user is required to enter a name for the experiment along with some basic information about the experimenter. The user can also supply information on the overall purpose and focus of the experiment, and references to relevant experiments, scientific publications or web resources.

Design Experiment

The experiment type can be selected here. A simple graphical representation of the experiment workflow is presented to the user. Configuration of this workflow depends on the experiment type, but in most cases, this involves browsing

Chiara Cirinnà and Maurizio Lunghi (edited by), *Cultural Heritage on line: Empowering users: an active role for user communities. Florence, 15th - 16th December 2009* ISBN 978-88-6453-184-7 (print) ISBN 978-88-6453-187-8 (online)

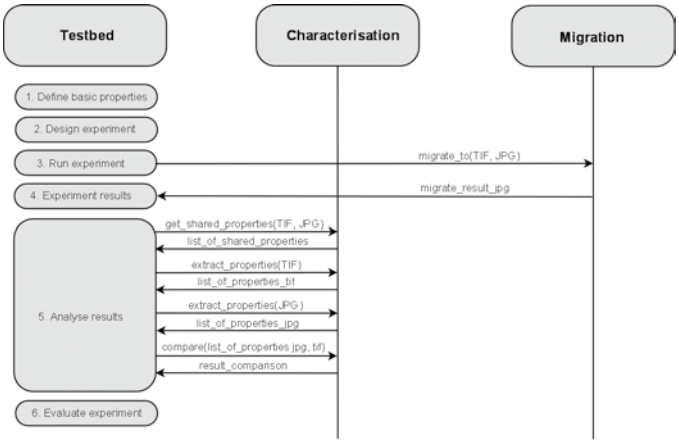


Figure 1: Example of a Planets Testbed experiment process

and selecting available services and selecting digital objects to experiment upon. The digital objects can be chosen from the data sets available in the Testbed or from content the user has uploaded.
Taking the example of the migration experiment, the workflow is configured by selecting a migration pathway, composed of the starting format TIF, the target format JPG, and a migration service (e.g. ImageMagik).

Run experiment
Once designed and configured, the experiment can be submitted for approval. At this point, the administrator in charge of the Planets Testbed is given an opportunity to prevent the experiment from being executed, for example if it is likely to put an unreasonable load on the server if executed at that time. Experiments that require only modest resources are automatically approved, and can be executed straight away.
Following approval, the user can initiate execution of the workflow. The Planets workflow execution engine then takes each digital object, and passes it through the specified chain of services.

Experiment results
In this step, the user can inspect the experiment result objects, overall success rates and basic performance statistics, e.g. whether all migration actions successfully created new digital objects. The user is also given the opportunity to re-run the experiment in order to collect additional data.

Analyse results
If characterisation tools are available for the digital objects which are part of the experiment, they can be used to analyse the properties of the digital objects. In our migration example, there are two digital objects, an input TIF file and the resulting JPG file which have different file format specific characteristics. Based on the common set of properties of these file formats which are determined by a Planets characterisation service, the values can then be automatically compared using the metrics that apply to the different properties.

Evaluate Experiment
The final step of an experiment allows the user to judge the overall performance of the preservation workflow. The experimenter can also provide a brief written report about the experiment's outcome. The result can then be more widely shared between Planets Testbed users, so that others can learn from the results or even setup an equivalent experiment in order to reproduce and verify the outcomes of other experimenters.

Planets Testbed experiment types
An experiment type defines the generic structure and data flow of an experiment, and there are many kinds of experiments to be explored other than the migration experiment outlined as an example above. In the following, we shortly describe the experiment types that exist so far.

- *Characterisation Experiments*
  A characterisation experiment allows for direct comparison of characterisation tools against each other or against a set of authoritative property values.

- *Validation Experiments*
  A validation experiment is used to test whether a digital object is well-formed and valid with respect to a particular format.

- *Emulation Experiments*
  Emulation generally refers to imitating a (usually obsolte) soft- and hardware environment within another (usually up to date) soft- and hardware environment. In the Testbed, an Emulation experiment creates an emulation session for a digital object which is then visualised the imitated soft- and hardware environment. By that way, the user can record how well the object is being rendered with respect to this specific environment.

- *Execute Plato preservation plan*
  "Plato" (see [1]) is one of the outcomes of the Planets project, and is a web based software for creating a preservation plan for preserving a specific collection or a part of a collection of digital objects. The concrete recommendation of the preservation plan ends up in an "executable preservation plan" which can then be evaluated by a corresponding Planets Testbed experiment.

It is to be expected that the existing experiment types do not cover all the requirements for the different experiment scenarios the long term preservation community might require. If an experiment does not fit with one of the existing experiment types, a new experiment type must be set up by a Testbed administrator contacted through the Testbed helpdesk (see end of section 3).

Sharing knowledge with the Planets Testbed community

The Planets Testbed is community software in the sense that it allows reviewing and even reproducing existing experiments by all community members. New experiments can reference existing ones and refine or give a statement on existing experiment results. In that way the community members contribute to a continuously growing and reliable knowledge base on digital preservation.
The main goal of the Planets Testbed in this aspect is to enable community members to share their research results amongst cultural heritage institutions all over Europe. The Planets Testbed acts as the central experimentation platform gathering knowledge about long term preservation topics in various dimensions: In the first place, an experiment can focus on performance and reliability of long term preservation services and the underlying software components themselves. Then, the annotated experiment datasets contain information about special cases (an extreme value for a file format specific parameter, for example) and important properties of digital objects. And finally, the Planets Testbed establishes a procedure to share meaningfully aggregated results with other Planets software, like Plato (see [1]), for example.

Knowledge about long term preservation services
A wide range of preservation services have been developed by the Planets project, and the Planets Testbed aims to make them available for public use. Each service is supplied with metadata describing the supported formats, migration pathways, the identity of the service creator, the location of the endpoint which makes the Planets service available and so on. The Planets Testbed makes it easy to explore this information which is continuously managed and maintained.

Knowledge about experiment datasets

Some experiment types require information about the data an experiment is based upon. The Planets Testbed integrates annotated datasets (corpora) in order to be able to check the output of a service against recorded metadata. As a simple example, if an identification tool is tested against an object of a known format (e.g. PDF file), the Planets Testbed can compare the embedded properties against the results from the identification service. This allows the scope and accuracy of identification tools to be closely examined. Similarly, validation services can be exercised using carefully constructed valid and invalid documents, testing the edge-cases of format specifications. For example, the Isartor test suite (http://www.pdfa.org/doku.php?id=pdfa:en:isartor) can be used to detect whether validation tools can spot PDFs that are invalid with respect to the PDF/A-1 (ISO 19005-1:2005) specification.

Contributing to the Planets-wide knowledge base

By standardising and sharing results, the Planets Testbed acts as a central point for accumulation and aggregation of data from many experiments and across institutional boundaries. From this rich dataset it should be possible to determine the robustness and performance of particular preservation tools and techniques in an objective manner.
The results are stored centrally and can be used as a basis for future development of a knowledge base.

The Public Planets Testbed

The Planets Testbed software will be made publicly available by the Planets project. A full installation requires all of the different preservation services to be installed, each of which may have different software dependencies and operating system requirements. The publicly available central Planets Testbed addresses this problem by providing as many tools and services as possible – pre-installed, configured and ready for testing. The Planets Testbed can be accessed using a web browser and allows interested parties to evaluate all the preservation services and strategies supported by Planets using their own data or benchmark content.
Additionally, it is possible to download and install individual Planets Testbed instances. The software installer makes it easy to deploy the Planets Testbed locally, but can only provide limited functionality out of the box.
The public Planets Testbed is available at http://testbed.planets-project.eu/testbed, hosted by HATII at the University of Glasgow. It is currently in beta release phase and selected external parties have accounts granted. The service will go completely public in beginning of 2010, but it is already possible to ask for an account at helpdesktb@planets-project.eu. Further information about the Planets Testbed, also about upcoming training workshops can be found on the Planets website.

Conclusions

The innovative aspects of the Planets Testbed are the ways in which experimental data is collected, analysed and shared. The Planets Testbed provides a single interface to a wide range of hardware and software benchmarking environments, so that data can be collected reliably and reproducibly.
The Planets Testbed is also building corpora of digital objects with well-known properties. These properties, in combination with a number of innovative Planets software technologies, allow for the outputs of preservation services to be analysed rigorously and automatically.
Finally, the Planets Testbed defines standard semantic structures to contain these results, permitting community-wide aggregation of experimental results and experiences using the tools and services needed for long-term preservation of the digital cultural heritage.
The Planets Testbed will be made available to the digital preservation community as a free service by beginning of 2010.

References

[1] Christoph Becker, Hannes Kulovits, Andreas Rauber, and Hans Hofman, Plato: a service oriented decision support system for preservation planning, JCDL '08:Proceedings of the 8th ACM/IEEE-CS joint conference on Digital libraries (New York, NY, USA), ACM, 2008, See http://doi.acm.org/10.1145/1378889. 1378954, pp. 367-370.

[2] Petra Helwig, Judith Rog, Caroline van Wijk, Eleonora Nicchiarelli, and Manfred Thaller, Test methods for testbed, Tech. report, 2007, See http://www.planets-project.eu/docs/reports/Planets_TB3-D2_MethodsForTesting.pdf.

Abstract
The Czech Republic has earned worldwide recognition for its remarkable results in the area of cultural heritage preservation. However, digitisation and digital preservation are significantly hindered by a lack of resources. This results in a relatively slow pace of digitisation. Furthermore, it leads to serious delays in dealing with, and solving, current digital preservation issues.
This paper explores a "National Digital Library" project, which has been accepted by the Ministry of Culture as a candidate for European funding under the Integrated Operational Programme. The National Library of the Czech Republic, along with the Moravian State Library in Brno, have prepared an ambitious project with two main goals – 1) to accelerate the digitisation; 2) to establish a trusted long-term preservation repository. 1.2 million documents should be digitised within the next 20 years. The most fragile documents should be digitised during the five-year project between 2010 and 2015.
The following paper reflects on the issues we have to face in preparation for this project. The mass digitising encourages the institutions which are involved to make a number of organisational changes. There are also a number of strategic decisions to be made (national/institutional digital preservation policy formulation, national bibliographic identifier scheme implementation). And there are also a number of technical tasks with the possibility of an enormous future impact (like decisions on what file formats and metadata formats to use for this mass digitising, how to choose LTP system software, how to include the data from previous projects etc.)
The paper concludes that planning large-scale digitising needs significant administrative, organizational and political preparation, which may be more overwhelming than the technical part of such a project. Involved institutions must be ready for a business change, well before the scanners produce the first pages.
Keywords: digital preservation; national policy; mass digitisation; EU project

Historical background

In the National Library of the Czech Republic (NLCR) digitisation started in the early 1990s and the webarchiving was launched in 2000. The first digitisation projects, financed from public grants of the Ministry of Culture of the Czech Republic (MCCR), were focused on old prints and manuscripts, later also on historical newspapers. Digitisation was then considered to be a way of preservation as it reformatted the documents in danger of deterioration. Later, naturally, the focus had changed, and turned more towards the end users needs. Our digital libraries tried to comply with existing international standards, and the digital content is being integrated into portals like TEL, EUROPEANA. Even though the Czech Republic is a small country, it has earned worldwide recognition for its long tradition and remarkable results in the area of culture heritage preservation: in 2005, the NLCR was awarded the first UNESCO/Jikji Memory of the World Prize for its contribution to the preservation and accessibility of our documentary heritage.
Until today the main projects have produced 80TB of data, and yet they cover only a small fraction of our national cultural heritage. With the current pace of digitising, we would be working for the next 300 years to make accessible the nation's cultural heritage in digital form. However, lack of sufficient funding slows down the digitisation process and leads to delays in dealing with digital preservation issues.

National Digital Library (NDL)

In 2005 MCCR with NLCR jointly prepared the National preservation policy for the traditional and electronic library documents until the year 2010. [1] Hereby proposed funding was 8 million Euro. The whole policy stayed only on the level of declaration, and the financing was never approved by any of the next Governments.
The image below explains the organization of digital preservation as described in the National preservation policy for the traditional and electronic library documents until the year 2010.

Chiara Cirinnà and Maurizio Lunghi (edited by), *Cultural Heritage on line: Empowering users: an active role for user communities. Florence, 15th - 16th December 2009* ISBN 978-88-6453-184-7 (print) ISBN 978-88-6453-187-8 (online)

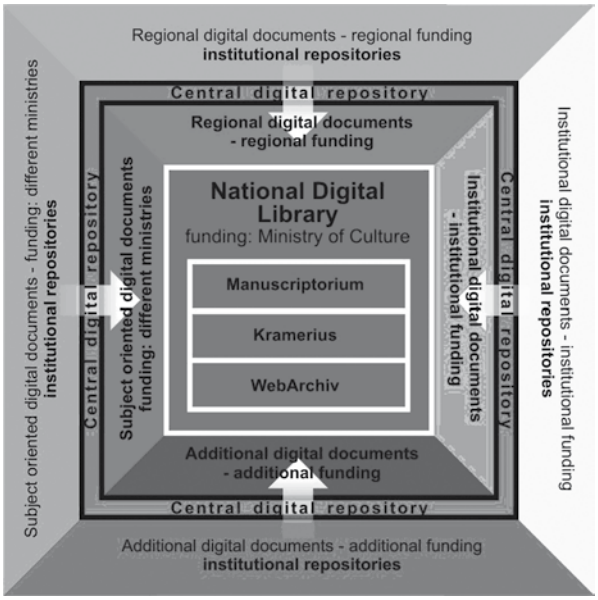


Figure 1: Project of the National Digital Library and its central repository

NDL forms the heart of the whole national system of culture heritage preservation. NDL contains the core of our national cultural heritage. These documents, digitised or born digital, are acquired, preserved and disseminated within three large national projects funded by the Ministry of Culture.

Manuscriptorium [2] is a system which gathers information on historical book resources, linked to a virtual library of digitised documents.

The Kramerius [3] project focuses on the preservation of and accessibility to 'modern' periodicals (from year 1801 onward), books and other documents in danger of acid paper degradation.

WebArchiv [4] is a digital archive of Czech web resources

Digital documents held by any Czech library, museum or archive can be selected to become part of the NDL. Digitising and preservation of these documents is to be funded by the MCCR. Digital data not selected for the NDL can also be deposited into the central repository, but their long-term preservation has to be funded from other resources. Other institutions may not be interested in depositing their data into the central repository. In such cases they have to secure their own financing, but their data can be integrated into the national access portals, provided that they comply with the metadata standards.

NDL operates in the broader context of a wider national digitising strategy of the MCCR, which also covers archival documents, museum collection, architectonic monuments, performing arts and media etc. However this wider strategy is currently only operating on a conceptual level, it's not a policy with proposed financing schemes.

NLCR AND DIGITAL PRESERVATION ISSUES: CURRENT STATE OF THE ART

NLCR is the key player in the area of long-term preservation in the country. Other libraries in the country rely on NLCR and wait for a solution they can follow. The nationwide standards for the digitisation projects are set by NLCR. Most of the projects use the same metadata schemas, the same access applications, the same or similar file formats.

The current state of digital preservation in NLCR is far from ideal. From the OAIS point of view NLCR only implements the archival storage module. Current installation, the Central Data Storage (CDS), is built on IBM products. Two IBM Systems Storage DS 4800 are installed, one in Klementinum and the second one in Hostivař data centre (18 km

distance). The Tivoli Storage Manager (TSM) is used for the back-up and archive services, together with an IBM tpe library. We also have archival back-up and an archiving strategy in place. Of course disaster recovery services are in place and replication between localities helps to protect data against physical destruction and human or software error. Some parts of OAIS functions are secured by number of applications used in the digitisation workflow or in the access applications. The ingest processes are limited to the hash function checks, consistency and completeness of the package and batch. Access applications use separated data storage and metadata database, and the long-term archival copies are seldom used. There is no metadata management module upon the archival data. Data arrives with several different formats of metadata, and there are often problems with identifying the archival documents, and providing links to the access copies and library catalogues. Many of the basic OAIS functions which enable digital object management and preservation planning are missing.
Access is secured by servers access applications, which have more or less reliable authorization and authentication mechanisms. Each of them has its own store of the user copies. International standards as DC and MARC21 (used as MARCXML) are used in most Czech libraries. Since 2007 we have used a structural, administrative and technical metadata scheme based on METS and PREMIS (the PREMISobject part).
NLCR have participated or NLCR currently participate actively in a number of European projects in the area of digitising (ENRICH, TEL+, TEL-ME-MOR, EDLnet) and digital preservation (DigitalPreservationEurope). Involvement in the DPE project was extremely important. In 2007 the central digital repository of the NLCR went through an audit based on the first generation of the DRAMBORA toolkit, and this helped the staff to realize the risk related with current underdeveloped preservation solutions.

Dramatic step forward: Towards the IOP project

The situation described above is destined to change very soon. The NLCR, along with the Moravian State Library in Brno (MLB), have prepared an ambitious project, which should be financed mainly from the Integrated Operational Program Smart Administration in 2010-2015. The feasibility study for the project was finished in September 2009, and the project calls should be open at the end of the same year. The project, called again "National Digital Library," has the following main goals:

- to accelerate digitisation (building two digitisation centres with robotic scanners in Prague and Brno for mass digitisation)
- to improve long-term preservation and access to digital objects (building a trusted and certified digital repository using two geographically separated localities - Prague and Brno, 180km from each other, purchasing digital preservation system software and tailoring it to the needs of the project).
- to secure wide dissemination of the national cultural heritage in digital form in a user friendly environment (using national aggregators and portals, possible also upgrading the technology of the national meta-search tools)

The digitising centers and the long-term preservation system have to be integrated into the existing infrastructure of the two participating libraries. Some further steps are needed to achieve permanent financial, technological and administrative sustainability of the created systems and of the digitised data and access to them.
The core of the Czech national cultural heritage (documents published in the country since 1801 + historical documents until 1800 stored in Czech libraries) form approximately 1.2 million documents, that is about 350 million pages. Many of the documents are printed on acid paper and/or are highly used and their digitisation is therefore very urgent. The projects infrastructure should allow digitising these 350 million pages within the next 20 years. The most fragile or highly used documents should be digitised during a five-year project itself between 2010 and 2015. The results of the project will be digitisation of 540.000 documents published since 1801, 20.000 documents published before 1800, archiving Czech web, all together producing about 1,5PB of data. The total budget of the project should be 27 million EUR (85% from European funding and 15% from co-funding).

The issues and decisions we have to face

The whole IOP project draws attention to a number of issues which need to be clarified before the mass digitisation will start. During the preparation of the project we have to face a number of organizational issues, and make a number

of strategic and technical decisions. It is clear now, that organizational and process changes are on the same level of importance as those of a technical nature.

Staffing and organization

First of all, even though the NLCR has been running a digitising line and some software and hardware infrastructure for many years, the processes of mass digitisation will require much more technology and staff on all levels. This will inevitably lead to a number of organizational and management changes. There was no "digital preservation department" in the NLCR before 2008, and no IT experts who would be able to coordinate and run the necessary environment, or survey the work of service companies. As well as more skilled staff members, changes to corporate culture are also necessary. The need for closer cooperation between different parts of the library became crucial in the process of project preparation. The communication channels still have to be improved between the IT departments, the new digital preservation team, the cataloguing department and the digitising team. Also, an experienced project manager will be needed to manage and control the entire project. All this means shifting the organization to a more business like culture, and to ensure there is a more cooperative environment inside the institution. Existing workflows in the library are currently undergoing reviews, so that we can locate the staff which could be relocated to digitising document selection and preparation. The preparation of the documents for digitising requires in many cases conversion of traditional catalogue records from scanned catalogues into Aleph database, or creating new catalogue records, de-duplicating the existing records etc.

This kind of project also needs the support of the top management of the library and the library funding body. Large digitisation still needs advocacy even inside the institution, as the library functions are multiple and the project will certainly affect the daily business in most departments.

Strategies, policies and politics

In a project of such extent we must consider the number of stakeholders' needs. The project is of potential interest to Authors' right holders associations, politicians, producers of HW and SW, similar projects in the country, and in other libraries. On the political level, the coordination with other Smart Administration projects would be necessary.

What more, NLCR has no clear policy statement on digital preservation, specifying the responsibilities and extent of preserved materials with clear long-term cost estimation and technical specifications and requirements. Even the Library Statute is missing a reference to this problem. NLCR prepared an internal draft of the general digital preservation policy of the institution, and suggested a review mechanism of this policy document. But this document has to be approved by the NLCR steering board, which has to recognize the budget and staffing. Besides, some more strategic decisions are to be made on the national level very quickly, especially about the URN:NBN identifier system implementation, and the national bibliographic number implementation. Both will be essential for the success of the mass digitising and preservation projects.

Technical issues

Naturally there are many more down to earth issues, which the project team had to face. Even little decisions are of potential large future impact. The first challenges we faced was to measure the amount of pages and documents we will have to digitise and the data amounts, in order to set the final sizes of the data repository and provide an estimation of the necessary scalability, setting the staffing needs of the project, measuring the economic efficiency of the project, etc. After completing this demanding process we realised that in many areas we only have sufficient data to make rough estimations, even though the founding agency would need a solid exact numbers.

We had to make decisions about the use of file formats, compression level, bit levels, metadata schemas, and other standards. We are determined to follow acknowledged standards field. [5] We expect to use mathematically lossless JPEG2000 for preservation master files and lossy JPEG2000 as a user copy. The OCR files in METS ALTO, and metadata files (XML METS with PREMIS, MIX, MODS or MARCXML). The webarchiving will move towards the WARC format from currently used ARCs. We have to design specific ingest workflows for various types of incoming digital documents, and decide how the existing digital data will enter the new digital preservation storage. This might be quite a time consuming and complex process of migration of old data and metadata into new formats, which would then be ingested into the new system. As we have about 7 millions of pages, this could take months.

At the highest production speed we expect to produce and archive more them 70 000 of pages a day on four robotic scanners in Prague and two scanners in Brno. This will also change the requirements on the access applications, and some will have to undergo technological reconstruction.
Since we have no real IT expert team capable of large scale programming and we do not expect to hire such a team in the future, we decided to search for a commercial solution for the long-term preservation system. Advantages seemed obvious – buying a "ready-to-go" system, which possibly already has some implementations in other libraries/archives. Even though the system will need some adjustments, it will fulfill most of the requirements for secure storage of the mass digitising production. Future support, development activities and upgrades will be done based on the requirements of more users of the system.
According to our present knowledge, there are only three commercial solutions available on the market now (SDB by Tessella, Rosetta by ExLibris, DIAS by IBM). After lot of workshops, corresponding with all providers, our team saw all of these systems running in the real implementations. In August 2009 we sent out an RFI to these three companies to design a complex solution for the central digital repository, which should guarantee the long-term digital preservation and dissemination of digital objects (including metadata). The RFI consisted of the description of the required system and list of requirements which we asked the companies to report back on. Their responses included cost estimations for the LTP system software, databases or other dependent software components, installation and setting costs, maintenance costs, required hardware infrastructure, and a prognosis of the running cost for the next ten years. We were also interested in receiving an estimation of time needed for LTP system implementation (from the contract to pilot phase and to productive phase).
The collected data were used in the feasibility study [6], with similar data from the producers of scanning workflows and scanners. The project was submitted in October 2009 and tenders covering different parts of the project are expected in spring 2010.

Conclusion

If it were only the technical decisions about file formats and metadata schemas, the preparation of this mass digitising project would be much simpler. However, the organizational and management aspects can influence the project results much more intensively then we could have expected, and may require much attention and workforce.

References

[1] NLCR, 2005. National preservation policy for the traditional and electronic library documents until the year 2010. Draft.
[2] http://www.manuscriptorium.com
[3] http://kramerius.nkp.cz/kramerius/Welcome.do?lang=en
[4] http://en.webarchiv.cz
[5] Florida Digital Archive, 2003. Recommended Data Formats for Preservation Purposes in the Florida Digital Archive. <http://www.fcla.edu/digitalArchive/pdfs/recFormats.pdf>.
[6] NLCR, 2009. Feasibility study for project National Digital Library.

Abstract
In the context of the eSciDoc project, the Max Planck Digital Library and the FIZ Karlsruhe are building an e-research environment for multi-disciplinary scientific research organizations. Based on the eSciDoc infrastructure, several solutions for the end-user will be developed and provided as open source software. One of them is ViRR (Virtueller Raum Reichsrecht), a solution to support collaborative and interdisciplinary research on text resources like manuscripts or books. A user-centred approach was applied to define necessary functionalities and adequate graphical user interfaces. ViRR provides several smaller flexible tools in one web interface for the creation and enrichment of metadata, for the modelling of the structure of a work and for the enhancement of the collection with related resources such as annotations and transcriptions. One of them is a configurable online editor for defining the structure of the digitized work in accordance with the structure of the original resource.
This paper will give an overview of the ViRR solution which was developed to support researchers from different backgrounds working together on text resources. Additionally, we will outline eSciDoc, the underlying infrastructure of the ViRR solution.
Keywords: eSciDoc, digitized text resources, collaborative workbench, online editor

Introduction

In the context of the eSciDoc project (http://www.escidoc.org) the Max Planck Digital Library (MPDL) has developed a web based solution for different user groups (researchers, librarians) to make their textual holdings online available. ViRR [1] enables the enrichment, dissemination and preservation of digitized cultural heritage like manuscripts or books. Its aim is mainly to support scholars in the humanities in the analysis and evaluation of text resources.
The MPDL is a scientific service unit within the Max Planck Society (MPG), which consists of about 80 institutes from various scientific disciplines, and therefore the development of services and solutions has to deal with requirements from diverse research contexts. During the development of the ViRR solution, the general approach was to start with specific requirements from a pilot community, and then identify generic services, which can be re-used by other disciplines. The aim is to develop a solution which can fulfil most of the diverse requirements of working with digitized text resources within the MPG.
The name ViRR derives from the content of the first collection, which consists of about 20.000 scans of legal artifacts from the period of the Holy Roman Empire provided by the Max Planck Institute for European Legal History (http://www.mpier.uni-frankfurt.de).

Working with Digitized Text Resources

Solutions, which support scholars in their work with digitized text resources, differ in their focus and quality as working instruments. The very basic level is the mere digital representation of a single text resource with basic browsing functions and without any sophisticated user management or re-use options.
A more enhanced level offers functionalities to intellectually enrich digitized text resources. Hereby, the scholars and librarians are able to uncover the “hidden” information, which cannot be provided by a mere digital representation. Some of these functionalities imply the capturing and enhancement of structural metadata and semantics, ideally in different standard formats like METS [2], MODS or TEI. Detailed information about the composition of a resource might be gathered, such as the pagination (logical and physical) or the structure of a work (see e.g. [3, 4]). Standardized interfaces support the re-use of this additional information in other contexts, such as library catalogues, aggregated viewing environments or mash-up services, and allow the integration of external knowledge bases, such as dictionaries or viewing tools.

Chiara Cirinnà and Maurizio Lunghi (edited by), *Cultural Heritage on line: Empowering users: an active role for user communities. Florence, 15th - 16th December 2009* ISBN 978-88-6453-184-7 (print) ISBN 978-88-6453-187-8 (online)

Having the resources and the related information on the web, the logical consequence is the support of collaborative scenarios from various disciplines, which might assist the creation of knowledge related to the artifacts [5]. The possibility to describe different entities of a resource on a semantic, lexical, etymological or pragmatic level, and to describe the relations of these entities to other resources such as annotations, transcriptions, images or dictionaries, enables a real workbench scenario for scholars in the humanities.
To provide a sustainable solution for supporting these different aspects of a workbench, we have chosen a gradual approach in the development: providing an online editor for the enrichment of structural information, at the same time developing robust content models, to enable future interlinking to other artifacts.

The eSciDoc Solution ViRR

The eSciDoc solution ViRR combines a set of tools (components) for publishing scientific content in one user interface. This includes the two key features, the electronic modeling and editing of the original source material (ViRR Editor, see Fig. 1) and its online representation in a digital library (ViRR Viewer, see Fig. 1). These features are often separated from each other and realized in different tools, so that data transformations between these tools become a necessary drawback. The integrated design of ViRR allows users to perform all working steps within one software solution.

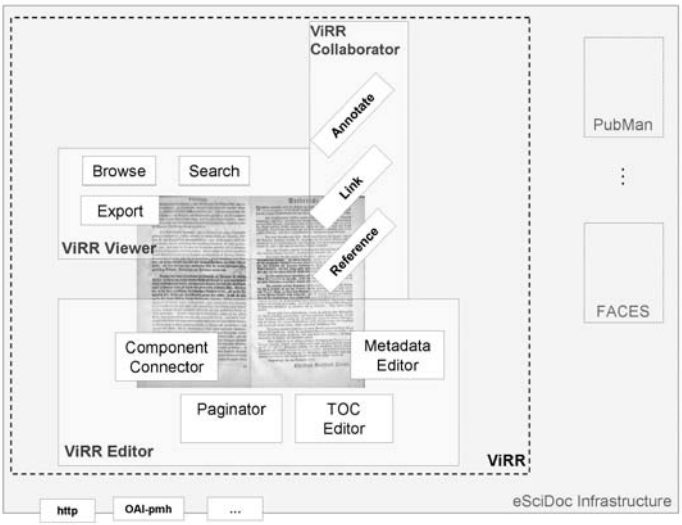


Figure 1: Overview of the different components of ViRR, embedded in the eSciDoc infrastructure

The core of ViRR is the online editor for the creation of electronic representations of cultural artifacts. While browsing through the scans (Fig. 2), several independent working steps are supported: the semi-automatic recording of the logical pagination next to the already available physical one (Fig. 3), the gathering of the structure via building a hierarchical tree based on different structural elements like, for example introduction, chapter or paragraph (ToC editor, Fig. 4) and the assignment of corresponding scans and descriptive metadata to these structural elements (metadata editor, Fig. 4). All of these working steps are presented in one complex, but flexible workspace. This design was chosen due to different user groups (e.g. librarians, scientists) with various working methods. It allows every user to configure the editor workspace based on his focus of work by providing relevant and hiding distractive information for each working step separately. Further on, all working steps can be performed in any order or can be mixed up depending on the individual needs of the user. Created data (structure, pagination, metadata) can be published online at any time during the editing process and therefore immediately be reused by other users.

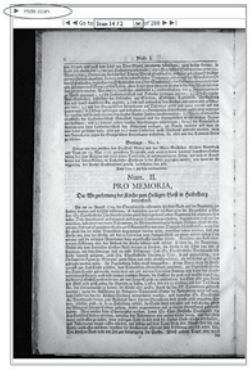


Figure 2: Parallel browsing during the editing process

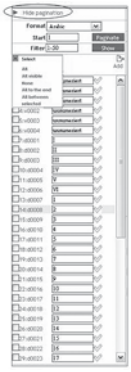


Figure 3: Paginator

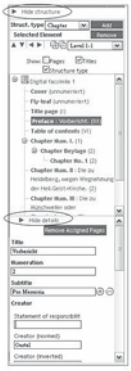


Figure 4: ToC and metadata editor

Within the ViRR Viewer, the content of the collection (multivolumes, volumes and monographs) is navigable via a browsing tree. Each work can be browsed separately in a configurable workspace where the user himself can decide whether he wants to see the bibliographic metadata, the logical structure in form of a table of contents or some parts of it, the scans, or a mixture of all of them. The offering of such customizable viewing sections provides each user an optimized environment to focus on his special interest.

ViRR Collaborative Aspects

In a next step, the ViRR solution will be enhanced with a new component, the ViRR Collaborator (as presented in Fig. 1), with the aim to improve the scientific value of the digitized collections by revealing hidden semantics and relations between various disciplines.

The provision of adequate collaboration tools is especially of interest when dealing with different research contexts: investigating textual aspects focus on certain details of a collection (e.g. transcriptions or the identification of text fragments) whereas studies on visual aspects focus on e.g. high resolution scans and referencing of image parts. Others might be interested in the collection as such by e.g. browsing through the scans and investigate the metadata. The challenge is to identify the generic functionalities for annotating and sharing, and to provide a working environment adaptable to the requirements of different holdings. Different collaboration tools can be applied like graphical annotations, e.g. by integrating the enhanced viewing environment DigiLib (http://digilib.berlios.de), or textual annotations. Further on, transcriptions of the original text corpora will be included to improve the semantically exploitation and retrieval of the digitized works. For easy creation and quality assurance of metadata, we will aim to integrate discipline specific authority data, either stored externally or provided by the eSciDoc service CoNE (Control of Named Entities [6]).

For supporting collaborative work around different collections we would like to enable users to invite others to co-work on a collection by assigning fine granular access rights to private content.

The eSciDoc Infrastructure

The collaborative refinements of the ViRR solution are mostly enabled by its underlying technical infrastructure. The eSciDoc infrastructure [7, 8] is designed as a service-oriented architecture. It is an open source joint development of the Max Planck Society and the FIZ Karlsruhe, funded by the German Federal Ministry of Education and Research (BMBF).

A service-oriented architecture fosters the reuse of existing services; therefore an eSciDoc service may be reused by other projects and institutions and become a building block within a broader e-Science infrastructure [9]. The data storage system for the eSciDoc infrastructure is based on the Fedora Commons platform (http://www.fedora-commons.org).

The eSciDoc content model primarily consists of two generic objects called item and container. An item object, in case of ViRR, is the digital representation of a cultural artifact (e.g. scanned page) and contains metadata (such as MAB, MODS) and optionally components (such as jpeg, pdf). A container object is an aggregation of objects (items or con-

tainers) such as a journal issue which aggregates several articles. Using this content model, ViRR specializes item and container objects into volume, multivolume, monograph, ToC, and scan (see Fig. 5).

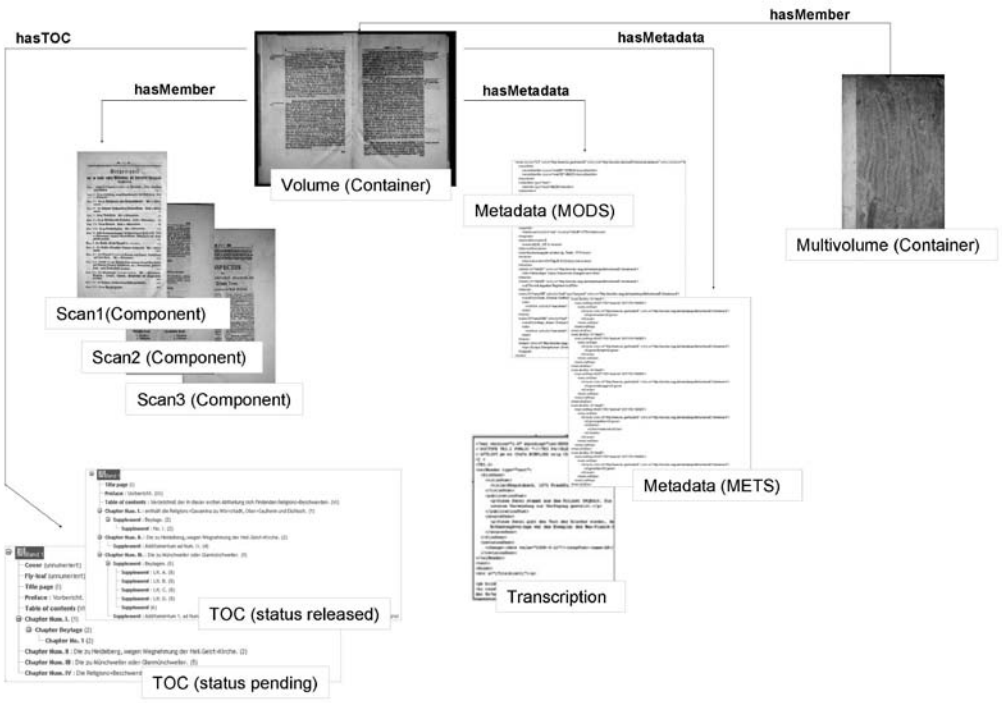


Figure 5: Digitized book content model

For example a digitized book is expressed in eSciDoc as a container, consisting of multiple items such as scans, transcriptions, and structural metadata. This container additionally holds the bibliographic metadata of the book (expressed in MODS). Additionally the generic data model of eSciDoc enables the integration and representation of data from diverse disciplines. The definition of new content models for other research data such as digitized journals or collections of images with discipline specific attributes can easily be integrated into the infrastructure, by defining a new content model with corresponding metadata profile.

As ViRR is fully embedded in the eSciDoc infrastructure it can profit from all existing eSciDoc services. Especially persistent identification (CNRI Handle or other), versioning, preservation (incl. PREMIS metadata) or the support of multiple metadata profiles (Dublin Core, MODS, custom profiles) would require, without the availability of eSciDoc, complex and time consuming development efforts for each new type of data.

eSciDoc is an open source project, setting a high priority in the implementation of standardized interfaces like oai-pmh, sword (http://www.swordapp.org) or RSS. Such an orientation fosters the integration of eSciDoc and eSciDoc-based solutions and their exploiting by other projects like the German national standardized viewing platform DFG Viewer (http://dfg-viewer.de). eSciDoc solutions are also evaluated in the context of other national or European initiatives like TextGrid (http://www.textgrid.de) or DARIAH (http://www.dariah.eu).

ViRR itself, besides offering functionality to process and disseminate data, provides as well services such as on-the-fly transformation of data. These can be used by other solutions, forming together an open accessible net of research data.

CONCLUSION

A range of requirements from different research disciplines exists for the handling of digitized cultural heritage on the web. Based on our experiences, this range can not be fulfilled by a monolithic software alone. One possibility to handle this range is to use an extensible infrastructure like eSciDoc, which focuses on standardization to support interoperability and therefore allows data exchange with services from other providers. So the data of eSciDoc solutions can be further re-used by external tools.

With the approach of using an underlying extensible infrastructure for the development of the ViRR solution, we are confident to fulfil most of the requirements arising from diverse disciplines concerning the work with digitized text resources, which is especially important in a heterogeneous research organization like the Max Planck Society.

REFERENCES

[1] http://colab.mpdl.mpg.de/mediawiki/ViRR:_Virtueller_Raum_Reichsrecht
[2] McDonough, J. P.: METS: standardized encoding for digital library objects. In: International Journal on Digital Libraries (2006)
[3] Gow, J., Buchanan, G., Warwick, C., and Rimmer, J.: Document Structure in Humanities Collections. http://www.cs.ucl.ac.uk/staff/J.Gow/papers/DocumentStructure.pdf (accessed 08 Sept. 2009).
[4] Bainbridge, D., Thompson, J., and Witten, I.H.: Assembling and enriching digital library collections. In: Proceedings of the 3rd ACM/IEEE-CS joint conference on Digital libraries. IEEE Computer Society, Washington (2003) 323-334.
[5] Hyman, M. D., and Renn, J.: Toward an Epistemic Web. Unpublished manuscript prepared for the Dahlem Workshop on Globalization of Knowledge and its Consequences. (2007)
[6] http://archimedes.fas.harvard.edu/mdh/epistemic-web.pdf (accessed 09 Oct. 2009)
[7] http://colab.mpdl.mpg.de/mediawiki/Service_for_Control_of_Named_Entities
[8] Bulatovic, N., Tschida, U., and Gros, A.: eSciDoc - a service infrastructure for management of Cultural Heritage content. In: Proceedings of the 14th International Conference on Virtual Systems and Multimedia, (Eds.) Ioannides, M.; Addison, A.; Georgopoulos, A.; Kalisperis, L. ARCHAEOLINGUA, Budapest (2008) 138-143.
[9] Dreyer, M., Bulatovic, N., Tschida, U., and Razum, M.: eSciDoc – a Scholarly Information and Communication Platform for the Max Planck Society. In: German e-Science Conference, Seq. No.: 315471.0 (2007).
[10] http://colab.mpdl.mpg.de/mediawiki/ESciDoc_SOA_AtGlance

Abstract
Supporting access to archived scientific publications, supplementary data, and multimedia objects as a basis for various types of reuse in scientific work processes and in publication processes is still an open issue in many ways. Reuse comprises, for instance, the subsequent verification of the content or its exploitation with a novel purpose. Retrieval approaches that factor in the versatile context of the archived data and documents can contribute to supporting reuse beyond traditional indexed based retrieval. The capturing of additional metadata during all life phases of digital objects before, during and after archival is a prerequisite to this approach. This paper motivates the usage of captured context data of digital objects for the purpose of enabling efficient reuse of preserved digital objects.
Keywords: OAIS, IR, context, scientific publishing

Introduction

An important goal of Digital Preservation (DP) is to enable the reuse of digital content. Reuse of digital content covers its subsequent verification and its exploitation with a novel purpose. Understanding the nature of the digital content and its origin supports information seekers in identifying relevant elements in archive collections and in interpreting them correctly. But the preservation of digital content, especially in the long term, covers periods of time, during which the nature of digital resources as well as their usage settings change [10]. As consumers cannot refer back to the creators, reuse of preserved digital objects depends on proper descriptions provided through the archive.
The SHAMAN (Sustaining Heritage Access through Multivalent ArchiviNg) project, co-funded by the European Commission under the seventh RTD Framework Programme, aims to develop a next generation digital preservation framework. The context model developed within the project provides an infrastructure-independent representation of the attributes associated with and (implied) relations between digital objects. Provided that the archive manages, preserves and makes available context data about digital objects, the SHAMAN context model is a potentially invaluable source for context-oriented retrieval on the archive holdings.
For SHAMAN context is not only defined by the discrete digital objects themselves, but also by the processes, in which they were created, ingested, accessed and reused. Processes are organized along phases within an Information Life Cycle Model. From an archive-centric perspective, each phase identifies one distinct stage in the life cycle of digital objects. Context comprises information about the preserved object itself, but also the relations between objects. Hence, capturing of contextual data is of great interest for enabling advanced retrieval in archival access, in addition to supporting preservation actions Through the preparation of context data the retrieval is not restricted on full text index, but could be opened to retrieval approaches on relations between objects. The retrieval results in this case are not necessarily preserved objects but could also be sets of contextual data.
This paper motivates the approach of context oriented Information Retrieval (IR), in an archive based life cycle with a focus on scientific publishing. This comprises current context oriented approaches in IR as well as the current approach towards a Context Model and an Information Life Cycle Phases Model in the SHAMAN project.

Contextual Data in the Domain of Scientific Publishing

Today, scientific publications are expected to be by origin born digital. They are presented and discussed at conferences and preserved over time in archives. Conferences take a prominent part in scientific research, because they are used to present works and ideas, to discuss new products, to determine trends, to socialize, and to initiate co-operation and collaboration. Conferences pay special attention to the assembling and the provision of scientific contributions.
Publications document the scientific contributions of the conference. Publications take various forms with individual strengths and weaknesses in distribution, storage, capacity and access capabilities. The abstract book documents the

Chiara Cirinnà and Maurizio Lunghi (edited by), *Cultural Heritage on line: Empowering users: an active role for user communities. Florence, 15th - 16th December 2009* ISBN 978-88-6453-184-7 (print) ISBN 978-88-6453-187-8 (online)

scientific contributions of the conference. Traditionally printed, abstract books nowadays are distributed as net publications. The conference web site allows for interactive structured access to the abstracts along date and time, type of presentation, topic and presenter, embedded in the overall scientific program of the conference.
Data collections incurred and managed in the course of one conference and/or consecutive editions of one conference features heterogeneous material with metadata and multitude of relationships. Entities in a collection comprise amongst others, abstracts, papers, posters, presentations, authors, sessions, and topics. Data and document collections comprise two general types: self-contained documents that can be considered complete and well-established (for example, presentation slides, posters, the printed conference abstract book), and the multitude of data, texts, images, and document parts gathered or produced in the course of the conference. This material includes amongst others, organizational data including conference participants, presenters, events, sessions, talks, and topics, as well as structured text information from conference contributions, especially abstracts and their tables and figures.
For a conference scientific contributions are accepted, indexed and re¬viewed. Speakers get invited, a scientific program is set-up, categorized and linked thematically.

INFORMATION LIFE CYCLE MODEL

Context data of digital objects evolve in different phases of existence. Context is guided by the processes in which the digital object is created, preserved, accessed and reused. Today, archives often depend on deriving metadata from the digital object obtained from the producer together with a minimum metadata set called-in by the archive. A good share of the imprint of the digital object gets lost during its transit into the archive. Opening-up the context of digital objects requires the capturing of context during all life phases of the digital object. Those life phases of a digital object are modeled in the archive-centric Information Life Cycle Model, depicted in Figure 1. The model distinguishes five relevant phases:

- Creation: new information comes into existence.
- Assembly: denotes the appraisal of objects relevant for archival and all processing and enrichment for compiling the complete information set to be sent into the future, meeting the presumed needs of the Designated Community. Assembly requires in-depth knowledge about the Designated Community in order to determine objects relevant for long-term preservation together with information about the object required for identification and reuse some time later in the future.
- Archival: addresses the life-time of the object inside the archive.
- Adoption: encompasses all processes by which accessed archival packages are unpacked, examined, adapted, transformed, integrated and displayed to be usable and understandable for the consumer. This includes also emulation activities if needed. The adoption phase might be regarded as a mediation phase, comprising transformations, aggregations, contextualisations, and other processes required for re-purposing data.

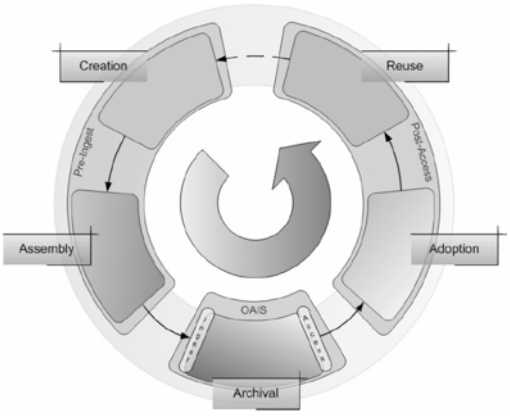


Figure 1: Information Life Cycle Phases

- Reuse: means the exploitation of information by the consumer. In particular, reuse may be for purposes other than those for which the Digital Object was originally created. Reuse of Digital Objects can lead to the Creation of other, novel Digital Objects. Reuse also may instigate the addition or updating of metadata about the Digital Object held in the archive. For example, annotation changes informational content and affects the relationships existing between the Object and other Digital Objects.

Context and its Representation

The pursued approach regarding context is Digital Preservation (DP) centric, following the Open Archival Information Systems (OAIS) Reference Model [1]. OAIS is a framework of terms and concepts providing a standardization of archival systems.

Context accords to the interrelated conditions in which something exists or occurs [3]. This expresses generally what all context definitions have in common. This statement implies for digital resource management, that the context of a digital object is complex, possibly containing concepts which are shared with other objects. This might be the process environment in which they are created, the associated actors, resources and information objects and also the preservation environment in which they are stored.

Furthermore, different domains and different scenarios have different requirements towards a context definition. Currently six content components are distinct in the context approach of the SHAMAN project: Document Context, Production and Reuse Context, Preservation System Context, Modeling Change Context, Social- and Enactment Context.

The most important context component for scientific publishing is the Production and Reuse Context (PRC). This context component corresponds to the producer and (anticipated) consumer environment, i.e. the respective designated communities creating and accessing digital objects. The creation environment includes the actors and resources involved, but also a formal representation of the organizational and technical processes carried out in the production of a digital object. To re-trace information paths, the representation of the production context has to be maintained during the transition from the production into the preservation environment. The reuse of preserved digital objects depends on a proper description of the significant properties and the associated domain-specific knowledge. This description allows for the efficient access and usage even from outside the designated community.

Especially in the PRC it is obvious that context is not only defined by the digital objects themselves, but it is also defined by the processes, in which they were created, preserved, accessed and reused. Domain-specific groundings provide interfaces to the relevant concepts and topics of the designated communities addressed, in addition to formalizations of the organizational structures involved, including associated role assignments. Concluding from this three distinct concepts are encountered, which are strongly involved in defining context. Those are:

- Domain: the concepts specific to the domain and their relations. For instance in the domain of scientific publishing: Abstract, Abstract Book, Presentation or Supplement.
- Enterprise: the structural layout of an organizational environment. For instance in the domain of scientific publishing: Affiliation, Persons or Roles.
- Process: the processes and their associated activities, including information about their implementations (service invocations): Submission, Indexing or Reviewing.

If context data should be preserved over time, a model for representation and organization of data is required. As a structured representation form of concepts and their relations, the usage of ontology is appropriate. An ontology represents concepts and their relations to one another. This could be seen as a formal model of a specific domain (see e.g. [5]). Ontologies are used to establish a common understanding about knowledge existing within a domain. One important aspect of ontologies is that they formally express the semantics of each element contained, enabling individuals and machines alike to access and process the knowledge represented. Rules and inference (or reasoning) mechanisms can be employed to derive new insights, i.e. making so far implicitly existing knowledge explicit.

The ontology used in SHAMAN is conceptually structured in the three sections Domain Ontology, Enterprise Ontology and Process Ontology. Those ontologies are consolidated through the ABC ontology [8], which was formally developed to model resources and their spatial, temporal, structural and semantic relationships.

CONTEXT-ORIENTED INFORMATION RETRIEVAL

Basing on the context notion and the representation of context as described in the previous sections, retrieval could be extended in two ways: firstly through the creation of an additional full-text index, containing the indexed context data and secondly through retrieval mechanisms on base of the relations between archived objects. Such relations between archived objects evolve through similar context attributes values. Those attribute values in the domain of scientific publishing are for instance the same author, the same conference, the same reviewer or common keywords. These data should be accessible through query, browsing with visualization support. The result of such a context oriented query is then not restricted on the archived objects; rather this could be a set of context data.

Context data in the domain of scientific publishing which can be expected to aid the retrieval of relevant publications for the purpose of scientific reuse are, for example

- Representation types such as abstract, presentation slides, poster or full paper;
- Embedding in the world of scientific discourse along citation nets, roles, interest and competence profiles of persons and organizations, and discussion threads;
- Implicit and explicit relationships to other documents like review reports and conference reports.

Those data could support, for instance, the retrieval of information for a state of the art research. Once a first relevant publication was found it could be used as the starting point to search for similar publications. Similarities according to publication context are, for example, but not limited to: publication origination from conferences with similar subject focus, by origination from the same conference, its conference sessions, its tutorials or its keynotes. Furthermore, it could be valuable to find publications with the same key words, publications which are referred to the source publication or the publications that refer to the source publication.

Different approaches for defining the concept context exist in IR. A user centric approach has been done for instance by Järvelin et al. in [6]. They stated that context is given through dependencies in time, place, history of interaction, task at hand and some other factors. Another approach towards a context definition in IR has been outlined by Cool et al. in [2]. They classify IR context in four different levels, namely: information environment, information seeking, IR interaction and the query level.

Some conceptual and implementation work on context based IR is already done. Melucci for instance presents in [9] a context model and the application of the model for ranking. Some context based IR support tools are implemented in Daffodil. This is an experimental system for IR and collaborative services in the field of higher education for the domain of computer science and others [4]. Daffodil comprises, for instance, an Author Net, which depicts relations among authors stored in a database and is used for ranking and the search for central actors in a set of documents or central actors for a specific author. Daffodil furthermore implements a Citation and Co-Author Browser, which are similar the Author Net, as well as an adaptive suggestion tool, which is based on the current situational user context [7].

But even if some particular solutions towards context oriented retrieval are implemented yet, the retrieval in preservation systems access lacks of offering a holistic model of digital object context for different domains and the preparation of context data for usage in retrieval.

For such a context oriented IR process it is essential to:

- define a holistic and adaptive context model, in order to serve the requirements of different domains
- provide mechanisms to capture relevant context data during the ingest phase
- prepare the context data in order to make them usable for retrieval
- offer an appropriate query- or browsing format in order to query the context data
- offer an appropriate way for presentation

The support of all those requirements is a task for future scientific work.

CONCLUSION

In this paper the advantage towards context oriented IR in an archive information life cycle is motivated. A context notion on basis of ontology is presented in order to model the context of preserved digital content. The ontology based representation provides valuable additional information for IR through the description of relations. By means of the

archive-centric information life cycle model, the important phases for capturing context are presented. The domain of scientific publishing was used to illustrate the usage of this retrieval approach.

References

[1] CCSDS. Reference Model for an Open Archival Information System (OAIS). Blue Book 1, Consultative Committee for Space Data Systems, January 2002. Recommendation for Space Data Systems Standards, adopted as ISO 14721:2003.
[2] Colleen Cool and Amanda Spink. Issues of context in information retrieval (IR): an introduction to the special issue. Information Processing Management, 38(5):605–611, 2002.
[3] Merriam-Webster Online Dictionary. context; cited 30.04.2009. "ONLINE" http://www.merriam-webster.com/dictionary/context.
[4] Norbert Fuhr, Claus-Peter Klas, André Schaefer, and Peter Mutschke. Daffodil: An Integrated Desktop for Supporting High-Level Search Activities in Federated Digital Libraries. In Research and Advanced Technology for Digital Libraries. 6th European Conference, ECDL 2002, pages 597–612. Springer, 2002.
[5] Nicola Guarino. Formal ontology and information systems. In Nicola Guarino, editor, Proceedings of the 1st International Conference on Formal Ontologies in Information Systems, FOIS'98, pages 3–15, 1998.
[6] P. Järvelin, K. & Ingwersen. Information seeking research needs extension towards tasks and technology. "ONLINE" http://InformationR.net/ir/10-1/paper212.html, 2004.
[7] Claus-Peter Klas, Sascha Kriewel, and Matthias Hemmje. An Experimental System for Adaptive Services in Information Retrieval. In Proceedings of the 2nd International Workshop on Adaptive Information Retrieval (AIR 2008), October 2008.
[8] Carl Lagoze and Jane Hunter. The ABC Ontology and Model. In Dublin Core Conference, pages 160–176, 2001.
[9] Massimo Melucci. A basis for information retrieval in context. ACM Trans. Inf. Syst., 26(3):1–41, June 2008.
[10] Ute Schwens and Hans Liegmann. Langzeitarchivierung digitaler Ressourcen. In Rainer Kuhlen, Thomas Seeger, and Dietmar Strauch, editors, Handbuch zur Einführung in die Informationswissenschaft und -praxis, volume 1 of Grundlagen der praktischen Information und Dokumentation, chapter D9, pages 567 – 570. München : Saur, 5., völlig neu gefasste Ausgabe. edition, 2004.

Abstract
In this paper we present the Italian initiative that involves relevant research institutions and national libraries, aimed at implementing an NBN Persistent Identifiers (PI) infrastructure based on a novel hardware/software architecture. This solution can be the base infrastructure towards the implementation of the European Global Resolver Service of PI.
The proposal is about a distributed and hierarchical approach for the management of an NBN namespace and illustrates assignment policies and identifier resolution strategies based on request forwarding mechanisms. Starting from the core motivations for the assignment of "persistent identifiers" to digital objects, this paper outlines a state of art in PI technologies, standards and initiatives, and illustrates other NBN implementations. The structure and goals of our initiative are described as well as the features already implemented in our system and the results of our testing activities. The paper ends with a proposal for the extension of this approach to the EU scenario.

Introduction
Stable and certified references to Internet resources are crucial for all the digital library applications, not only to identify a resource in a trustable and certified way, but also to guarantee continuous access to it over time. Current initiatives like the European Digital Library (EDL) [1] and Europeana [2], clearly show the need for a certified and stable digital resource reference mechanism in the cultural and scientific domains. The lack of confidence in digital resource reliability hinders the use of the Digital Library as a platform for preservation, research, citation and dissemination of digital contents [15]. A trustworthy solution is to associate to any digital resource of interest a PI that certifies its authenticity and ensures its long term accessibility. Actually some technological proposals are available [24], but the current scenario shows that we can't expect/impose a unique PI technology or only one central registry for the entire world. Moreover, different user communities do not commonly agree about the granularity of what an identifier should point to.
In the Library domain the National Bibliography Number (NBN – RFC3188) has been defined and is currently promoted by the CENL. This standard identifier format assumes that the national libraries are responsible for the national name registers. The first implementations of NBN registers in Europe are available at the German and Swedish National Libraries.
In Italy we are currently developing a novel NBN architecture with a strong participation of the scientific community, leaded by the National Research Council (CNR) through its Central Library and ITC Service. We have designed a hierarchical distributed system, similar to the DNS , in order to overcome the criticalities of a centralised system and to reduce the high management costs implied by a unique resolution service. Before describing our system in detail, we will provide in the following sections an overview of available PI technologies.

Persistent Identifier standards
The association of a PI to a digital resource can be used to certify its content authenticity, provenance, managing rights, and to provide an actual locator. The only guarantee of the actual persistence of identifier systems is the commitment shown by the organizations that assign, manage, and resolve the identifiers [25], [26].
At present some technological solutions are available but no general agreement has been reached among the different user communities. We provide in the following a brief description for the most widely diffused ones. Only the NBN [3] standard will be described in details in the next section.
The Document Object Identifier system (DOI [11]) is a business-oriented solution widely adopted by the publishing industry, which provides administrative tools and a Digital Right Management System (DRM).
Archival Resource Key (ARK [10]) is an URL-based persistent identification standard, which provides peculiar functionalities that are not featured by the other PI schemata, e.g., the capability of separating the univocal identifier assigned to a resource from the potentially multiple addresses that may act as a proxy to the final resource.

Chiara Cirinnà and Maurizio Lunghi (edited by), *Cultural Heritage on line: Empowering users: an active role for user communities. Florence, 15th - 16th December 2009* ISBN 978-88-6453-184-7 (print) ISBN 978-88-6453-187-8 (online)

The Handle System ([12], [26], [27]) is a technology specification for assigning, managing, and resolving persistent identifiers for digital objects and other resources on the Internet. The protocols specified enable a distributed computer system to store identifiers (names, or handles) of digital resources and resolve those handles into the information necessary to locate, access, and otherwise make use of the resources. That information can be changed as needed to reflect the current state and/or location of the identified resource without changing the handle.
Finally, the Persistent URL (PURL [13]) is simply a redirect-table of URLs and it's up to the system-manager to implement policies for authenticity, rights, trustability, while the Library of Congress Control Number (LCCN [14]) is the a persistent identifier system with an associated permanent URL service (the LCCN permanent service), which is similar to PURL but with a reliable policy regarding identifier trustability and stability.
This overview shows that it is not viable to impose a unique PI technology and that the success of the solution is related to the credibility of the institution that promotes it. Moreover the granularity of the objects that the persistent identifiers need to be assigned to is widely different in each user application sector.

NBN overview

The National Bibliography Number (NBN) [3] is a URN namespace under the responsibility of National Libraries. The NBN namespace, as a Namespace Identifier (NID), has been registered and adopted by the Nordic Metadata Projects upon request of the CDNL and CENL. Unlike URLs, URNs are not directly actionable (browsers generally do not know what to do with a URN), because they have no associated global infrastructure that enables resolution (such as the DNS supporting URL). Although several implementations have been made, each proposing its own means for resolution through the use of plug-ins or proxy servers, an infrastructure that enables large-scale resolution has not been implemented. Moreover each URN name-domain is isolated from other systems and, in particular, the resolution service is specific (and different) for each domain.
Each National Library uses its own NBN string, independently and separately implemented by individual systems, with no coordination with other national libraries and no commonly agreed formats. In fact, several national libraries have developed their own NBN systems for national and international research projects; several implementations are currently in use, each with different metadata descriptions or granularity levels.
Examples are the DIVA project [16], EPICUR [18], and ARK at National Library of France [17].
There are some important initiatives at European level like the TEL project that it is in the process of implementing a unique system based on NBN namespace within the European Digital Library (EDL). The adoption of NBN identifiers is needed for implementing the 'National Libraries Resolver Discovery Service' as described in the CENL Task Force on Persistent Identifiers report [19].
In our opinion NBN is a credible candidate technology for an international and open PI infrastructure, mainly because it is based on an open standard and supports the distribution of the responsibility for the different subnamespaces, thus allowing the single institutions to keep control over the persistent identifiers assigned to their resources.

The NBN initiative in Italy

The project for the development of an Italian NBN register/resolver started in 2007 as a collaboration between "Fondazione Rinascimento Digitale" (FRD), the National Library in Florence (BNCF), the University of Milan (UNIMI) and "Consorzio Interuniversitario Lombardo per l'elaborazione automatica" (CILEA). After one year of work a first prototype demonstrating the viability of the hierarchical approach was released. The prototype leveraged some features of DSpace and Ark and provided a basic PHP web interface for library operators and final users. The hierarchy was limited to a maximum of two levels.
The second and current phase of the Italian NBN initiative is based on a different partnership involving Agenzia Spaziale Italiana (ASI), Consiglio Nazionale delle Ricerche (CNR), Biblioteca Nazionale Centrale di Firenze (BNCF), Biblioteca Nazionale Centrale di Roma (BNCR), Istituto Centrale per il Catalogo Unico (ICCU), Fondazione Rinascimento Digitale (FRD) and Università di Milano (UniMi).
The Italian National Research Council (CNR) developed a second prototype based on Java Enterprise technologies and web 2.0 user interface, which eliminated the need for DSpace and Ark and the two-level limit and introduced new features. CNR and FRD hold property rights of the software and will release it as opens source under the terms of

EUPL license. In order to encourage its adoption by other national registers a supporting community will be established.
The results are available as an installable software; future objectives have been defined in order to extend functionality and integrate the system within an international infrastructure. To this end, the Italian group is currently establishing international collaborations.
In the following we provide a description of objectives, governing structure and licensing policy defined for the Italian initiative.
The initiative aims at:

1. creating a national stable, trustable and certified register of digital objects to be adopted by cultural and educational institutions;
2. allowing an easier and wider access to the digital resources produced by Italian cultural institutions, including material digitised or not yet published;
3. encouraging the adoption of long term preservation policies by making service costs and responsibilities more sustainable, while preserving the institutional workflow of digital publishing procedures;
4. implementing a new service based on URN, similar to other national systems but with a more advanced architecture in order to achieve distribution of responsibility for name management;
5. extending as much as possible the adoption of the NBN technology and the user network in Italy;
6. developing an inter-domain resolution service (e.g., NBN Italy and NBN Germany, or NBN Italy and DOI) with a common meta-data format and a user-friendly interface (pre-condition for global resolver);
7. creating some redundant mechanisms both for duplication of name-registers and in some cases also for the digital resources themselves;
8. overcoming the limitation imposed by a centralised system and distributing the high management costs implied by a unique resolution service, while preserving the authoritative control.

In order to define organization and policies for the Italian register, a governing board has been established, where BNCF, BNCR , CNR, FRD, ICCU are represented. The governing board defines the top-level structure of the Italian NBN domain hierarchy and the policies for overall infrastructure management, sub-domain creation/removal and PI assignment.

THE DISTRIBUTED ARCHITECTURE APPROACH

The proposed architecture, starting from [22], [23] and taking into account the URN standard requirements as [20], [21], introduces some elements of flexibility and additional features as shown in [29]. At the highest level there is a root node, which is responsible for the top-level domain (IT in our case). The root node delegates the responsibility for the different second-level domains (e.g.: IT:UR for University and Research, etc.) to second-level naming authorities. Sub-domain responsibility can be further delegated using a virtually unlimited number of sub-levels (eg.: IT:UR:CNR, IT:UR:UNIMI, etc.). At the bottom of this hierarchy there are the leaf nodes, which are the only ones that harvest publication metadata from the actual repositories and assign unique identifiers to digital objects.
Each agency adheres to the policy defined by the parent node and consistently defines the policies its child nodes must adhere to.
It is easy to see that this hierarchical multi-level distributed approach implies that the responsibility of PI generation and resolution can be recursively delegated to lower level sub-naming authorities, each managing a portion of the domain name space. Given the similarity of the addressed problems, some ideas have been borrowed from the DNS service.
Within our architecture each node harvests PI information from its child nodes and it is able to directly resolve all identifiers belonging to its domain and sub-domains. Besides, it can query other nodes to resolve NBN identifiers not belonging to its domain. This implies that every node can resolve every NBN item generated within the NBN:IT sub-namespace, either by looking up its own tables or by querying other nodes. In the latter case the query result is cached locally in order to speed up subsequent interrogations regarding the same identifier.
This redundancy of service access points and information storage locations increases the reliability of the whole infrastructure by eliminating single points of failure. Besides, reliability increases as the number of joining institutions grows up.

In our opinion a distributed architecture also increases scalability and performance, while maintaining unaltered the publishing workflows defined for the different repositories.

Policy

The trustability and reliability of an NBN distributed infrastructure can be guaranteed only by defining and enforcing effective policies. To this end the Italian NBN governing board is going to release a general policy that will have to be signed by all the participating agencies.

We have performed an initial analysis to detect problems and issues that the policy should address. In our opinion each agency should satisfy some requirements, which are both technical and organisational, and should commit in respecting some guidelines.

Organisational requirements

Each participating agency should indicate an administrative reference person, who is responsible for policy compliance as regards the registration and resolving procedures as well as for the relationships with the upper and lower level agencies, and a technical reference person, who is responsible for the hardware, software and network infrastructure.

Technical requirements

The hardware hosting an NBN register/resolver should be housed in a managed hosting infrastructure, with uninterruptable power supply and high-speed network connection. An agency that does not have an internal server farm may outsource hosting services to an external provider, which fulfils the technical requirements.

The hardware architecture should be redundant in order to guarantee no single point of failure.

In our opinion it would be also useful to identify and monitor some simple service level indicators, such as service response time and up time, and define thresholds that each agency should respect. Each domain maintainer could monitor its child sub-domains and notify them service level violations. The policy should also define how violations should be dealt with.

Guidelines

The policy should define rules for:

1) generating well-formed PIs;
2) identifying the digital resources which “deserve” a PI;
3) identifying resource granularity for PI assignment (paper, paper section, book, book chapter, etc.)
4) auditing repositories in order to assess their weaknesses and their strengths (the Drambora toolkit may help in this area).

Testing activities

After developing a first working prototype, collaborations have been established with several research institutions in order to create a community where final users and software developers are both represented. Several institutions are already involved in user requirement definition or have declared their availability to join the NBN network. These institutions are: the University & Research Group (ISS, INAF, INFN, INGV, ASI, ENEA, INOA, APAT, University of Pisa, University of Rome ‘Sapienza’, the University of Florence, the Florence University Press, University of Milan, i.e.).

A first testbed has been deployed where users can execute test cases and provide feedback to the developers in terms of bug/defect notifications, change or enhancement requests and new requirements. On the other hand the developers perform technical tests to evaluate performance, scalability and reliability of the infrastructure and implement what needed to satisfy user indications.

The testbed is configured as follows:

a) central node at BNCF, responsible for the Italian sub-domain (NBN:IT),
b) a second level inner node at CNR, responsible for the “University and Research” sub-domain (NBN:IT:UR),
c) a second level leaf node at FRD, responsible for the local NBN:IT:FRD sub-domain,
d) a third level leaf node at UNIMI, responsible for the local NBN:IT:UR:UNIMI sub-domain,
e) a third level leaf node at CNR, responsible for the local NBN:IT:UR:CNR sub-domain.

The second level CNR inner node (NBN:IT:UR) aims at implementing the University and Research National Registry. It currently aggregates the records generated by the UNIMI and CNR leaf nodes for the resources stored in their local repositories. The FRD node generates NBNs for resources stored in a local Dspace repository. A first set of tests has been performed to verify functionalities and behaviour in a distributed environment using different metadata sets. Performance was not the main focus in this phase and this is the reason why the servers used to set up the infrastructure are neither particularly powerful nor up to date.
First feedbacks from users are positive as regards registering and resolving functionalities. The system harvests resources, assigns NBNs and provides access to metadata and documents as expected. As regards duplicate discovery via hash comparisons, it has been pointed out that this mechanism works only if the compared files are identical, but fails even if they differ for a single bit. It has also been remarked that currently it is not possible to represent within the identifier the "part of" relation between two digital objects. This means that if we want to assign identifiers both to an entire document and to parts of it (e.g. a picture) there is currently no commonly agreed way to represent this inclusion relation in the final part of the persistent identifier. Finally, the need for higher-level services has been expressed by several parties, first of all the possibility of producing reports about the number of publications deposited in a sub domain within a certain period. This problem is tightly related to the duplicate detection one. If the latter is not solved, resource accounting statistics may be affected by errors whose impact cannot be estimated at the moment.

TOWARDS THE EUROPEAN RESOLUTION SERVICE

In this paper we have described a new software application for a distributed and hierarchical NBN register/resolver infrastructure. The main technical problems pointed out so far pertain to the identifier uniqueness guarantee. The proposed solution of using MD5 hash codes partly resolves this issue but poses performance problems and does not cover cases where the same content is represented in different formats. A more comprehensive solution will probably involve the comparison of a strictly defined set of metadata. This means that strict rules and clear responsibilities must be defined as regards data entry in the digital libraries.
From a political point of view the short-term objective is to enlarge the group of supporting institutions in order to create a first nucleus of a credible NBN national infrastructure. On a larger scale, CNR and FRD participate to the PersID project, funded by the Knowledge Exchange consortium and the SURF foundation, and aimed at developing a European Global Resolver. The adoption of our software as top-level node manager will be taken into consideration in the following months.
In our opinion it is also important to identify high-level value-added services (such as digital resource accounting) that could be built on top of the infrastructure. This would probably favour the diffusion of NBN persistent identifiers.
From the technical point of view the next steps will include performance testing and tuning, in order to define the hardware requirements for a production infrastructure that would guarantee the necessary service levels.
The testbed will be enlarged in order to include a leaf node installed at the University of Bologna, which will harvest records from the "Magazzini digitali" project repository. The goal of this project is to enable the BNCF digital library to harvest doctoral thesis from the University of Bologna Eprints repository, in order to accomplish their legal deposit. In this case the resources already have an NBN name. A new NBN record will be created in our registry using the existing identifier, which will be associated to the new URL assigned by legal deposit at BNCF.
A research group has also been established to thoroughly examine the duplication problem and its possible solutions. In this field hash codes different from MD5 could provide better performance with respect to comparison operations. The same group will also address the problem of the "part of" relation representation. Finally, we are going to investigate ways to establish permanent and reliable connections between NBNs and other persistent identifiers such as DOI, which would favour the implementation of a multi-standard global resolver.

REFERENCES

[9] European Digital Library http://www.theeuropeanlibrary.org/portal/organisation/cooperation/archive/edlproject/
[10] Europeana www.europeana.eu/
[11] IETF RFC 3188 Using National Bibliography Numbers as Uniform Resource Names http://tools.ietf.org/html/rfc3188
[12] IETF RFC 2141 URN Syntax http://tools.ietf.org/html/rfc2141

[13] C. Lagoze and H. V. de Sompel. The Open Archives Initiative Protocol for Metadata Harvesting, version 2.0. Technical report, Open Archives Initiative, 2002. http://www.openarchives.org/OAI/openarchivesprotocol.html.
[14] Dublin Core Metadata Initiative. Dublin Core Metadata Element Set, Version 1.1. http://dublincore.org/documents/dces/.
[15] MPEG-21, Information Technology, Multimedia Framework, "Part 2: Digital Item Declaration," ISO/IEC 21000-2:2003, March 2003.
[16] METS, <http://www.loc.gov/standards/mets/
[17] Herbert Van de Sompel et al. Resource Harvesting within the OAI-PMH Framework D-Lib Magazine December 2004 Volume 10 Number 12 ISSN 1082-9873
[18] J. Kunze. The ARK Persistent Identifier Scheme. Internet Draft, 2007 http://tools.ietf.org/html/draft-kunze-ark-14.
[19] Norman Paskin. Digital Object Identifiers. Inf. Serv. Use, 22(2-3):97–112, 2002 http://www.doi.org
[20] Sam X. Sun. Internationalization of the Handle System - A persistent Global Name Service. 1998. http://citeseer.ist.psu.edu/sun98internationalization.html. www.handle.net
[21] Persistent URLhttp://purl.oclc.org
[22] Library of Congress Control Number http://www.loc.gov/marc/lccn.html
[23] David Giaretta, Issue 1, Volume 2 | 2007 The CASPAR Approach to Digital Preservation The International Journal of Digital Curation
[24] Andersson, Stefan; Hansson, Peter; Klosa, Uwe; Muller, Eva; Siira, Erik Using XML for Long-term Preservation: Experiences from the DiVA Project
[25] Bermes, Emmanuelle, International Preservation News, Vol 40 December 2006, pp 23-26 Persistent Identifiers for Digital Resources: The experience of the National Library of France http://www.ifla.org/VI/4/news/ipnn40.pdf
[26] Kathrin Schroeder. Persistent Identification for the Permanent Referencing of Digital Resources - The Activities of the EPICUR Project Enhanced Uniform Resource Name URN Management at Die Deutsche Bibliothek. The Serials Librarian, 49:75–87(13), 5 January 2006.
[27] CENL Task Force on Persistent Identifiers, Report 2007 http://www.nlib.ee/cenl/docs/CENL_Taskforce_PI_Report_2006.pdf.
[28] Sollins, Karen Architectural Principles of Uniform Resource Name Resolution (RFC 2276) http://www.ietf.org/rfc/rfc2276.txt
[29] Masinter, Larry; Sollins, Karen Functional Requirements for Uniform Resource Names (RFC 1737) http://www.ietf.org/rfc/rfc1737.txt
[30] E. Bellini, M.Lunghi, E. Damiani, C. Fugazza, 2008, Semantics-aware Resolution of Multi-part Persistent Identifiers, WCKS 2008 conference.
[31] E. Bellini , C. Cirinna, M. Lunghi, E. Damiani, C. Fugazza, 2008 Persistent Identifiers distributed system for cultural heritage digital objects, IPRES2008 conference
[32] H.-W. Hilse, J. Kothe Implementing Persistent Identifiers: overview of concepts, guidelines and recommendations, 2006, ix+57 pp. 90-6984-508-3 http://www.knaw.nl/ecpa/publ/pdf/2732.pdf
[33] DCC Workshop on Persistent Identifiers, 30 June – 1 July 2005 Wolfson Medical Building, University of Glasgow, http://www.dcc.ac.uk/events/pi-2005/
[34] ERPANET workshop Persistent Identifers, Thursday 17th - Friday 18th June 2004-University College Cork, Cork, Ireland , http://www.erpanet.org/events/2004/cork/index.php
[35] Handle System website, http://www.handle.net/
[36] Wikipedia Handle page, http://en.wikipedia.org/wiki/Handle_System
[37] E. Bellini , C. Cirinnà, M. Lunghi, R. Puccinelli, M. Lancia, B. Sebastiani, M. Saccone, M. Spasiano - Persistent identifier distributed system for digital libraries - IFLA 2009 Conference – Milan

Abstract
The April 16 Archive (http://www.april16archive.org) at the Center for Digital Discourse and Culture is a memory bank of user contributed digital artifacts relating to the event, the April 16 Tragedy at Virginia Tech. A memory bank collects things that people contribute to it; usually digital originals, related to something worth remembering. These digital memorabilia form an esoteric collection that before its collection would have become ephemeral and likely lost (Goff, 2008). However, as collected, they should, as they are tied directly to the act of contribution and the more significant relations beyond that between the individual and the event, have significant social and emotional ties to the contributors and the community. This paper argues that those ties are fading. The April 16 Archive, as an event driven memory bank, originated from a passionate and committed community of users who shared the emotional and social attachment surrounding the event. The paper describes the tensions in the development and maintenance of the archive as the various communities have, through time, grown farther from the event, placing it further into their communal memories. In doing this, I hope to provide insights into the problems that develops in event driven digital archives as its communities grow apart. I also hope to share some of our experiences in developing and maintaining an event driven archive using web 2.0 oriented software.
Keywords: Memory Bank, Audiences, Users, Archives, Web 2.0

Introduction
As I sit in my office watching the tail of the web server logs scroll through my terminal window for a few minutes considering how I should start this paper, I am struck by how rarely the topic of this paper, the April16archive.org website is appearing in those logs. This observation is in part the basis for this paper. Event-driven archives, like the April 16th Archive struggle to maintain users overtime as the memories of the event fades, even if the effects of the event do not. On April 16th, 2007, Seung-Hui Cho, shot and killed 32 people and wounded many other faculty, staff, and students Virginia Polytechnic Institute and State University (Virginia Tech) . He left wounds both physical and emotional. While I was not on campus that year, I was still affiliated with Virginia Tech, as I am today, and like many I was overwhelmingly concerned for the safety and wellbeing of my colleagues at Virginia Tech. This event was a shock across the global university system and has had broad ranging effects from changing how universities deal with mentally ill students, to how universities manage security, and how we communicate with students, families, faculty, staff, and the greater audience. In short, this event was tragic and transformative, changing the ways universities and higher education operates in many ways.
Within a few days of the event, my colleague Brent Jesiek, who was managing the Center for Digital Discourse and Culture (CDDC) and now is an Assistant Professor at Purdue University, met with colleagues on campus from departments in the social science and humanities and they discussed options for preserving elements of the event that might be overlooked or uncollected elsewhere. Their idea was to move beyond the archival mission and into the memory mission, even toward a memorial mission, which would enable more personal and shared narratives, such as podcasts, blog posts and similar media to be captured as part of a memory bank like occurred for Hurricane Katrina and September 11th with their respective memory banks . Unlike a normal archive, the idea for the April16archive was to be more expansive:
"This project contributes to the ongoing efforts of historians and archivists to preserve the record of this event by collecting first-hand accounts, on-scene images, blog postings, and podcasts. It is our sincere hope that this site can contribute to a collective process of healing, especially as those affected by this tragedy tell their stories in their own words. The April 16 Archive runs on Omeka, a "digital memory bank" platform that uses the Internet to preserve the past and make memories available to a wide audience for generations to come". (http://april16archive.org/about)

Contacting colleagues George Mason University's Center for New Media and History (CNMH), Prof. Jesiek discussed the memory banks for hurricane Katrina and Sept. 11 and inquired as to the nature of the software used. The CNMH had software in development that was to become Omeka that they contributed for the CDDC to use on this project. Omeka was still in development at that stage, but one of the CNMH developers had recently graduated from Virginia Tech's history program, between his efforts and the rest of the CNMH staff, within a few days, the software to launch the April16tharchive was in Prof. Jesiek's email inbox. Eight days after the tragic events of April 16, on April 24th, the first objects from the general public began to be donated to the archive. On April 30th, the College of Liberal Arts and Human Sciences did a formal press release announcing this new memory bank to the [1].
Omeka is a memory bank application that allows the collection and discussion of digital collections in a web 2.0 environment. It is the platform that CDDC uses to host the April16archive today. Omeka is an object-oriented php/mysql web-based software application that can easily be installed and used for a variety of purposes. One such purpose is the memory bank. A memory bank attempts to collect and preserve memories--contributed objects such as images, videos, even word documents. Almost all the material contributed are ephemera, which might not be otherwise preserved were they not found, made, or even constructed to be preserved by their contributors.

EPHEMERA, EVENTS, MEMORIES AND AUDIENCES

In the April16archive, we focus on one event and this has dramatic effects in the two plus years that we have been in existence. As an event-driven archive--an archive that documents a temporal event that occurred at a certain time--the April16archive faces the challenge of sustaining a group of users or even a strong audience. In the beginning of the archive this issue was not foreseen. In November of 2007, we presented at the 48th annual Rare Books and Manuscripts section of the ACRL's Collecting for Contemporary Events seminar at Johns Hopkins University, where reported that at that time we had almost 1000 items in our digital collection, and today we have just over 1200, though we have secondary collections like the April 16th Archive Frontpages collection, which adds to that number. In addition, in the last year we have had few, actually very few, contributions to the archive that were not created in house. The drop off in contributions was matched by an increase in spam, which we eventually controlled. However, new contributions are occurring in the order of 1 or 2 every 3-5 months. In short, it appears as if, as one might expect that the memory of the event has faded and with that fading of memory, there has been a fading of the number of contributions.
Contributions to the archive are all similar to the image below:

Screenshot of archive material donated under Creative Commons License

In this image we can see all of the user-contributed content of a contribution to the archive. The author provides title, description, license information (if appropriate), tags and the content itself, which in this case is a large picture of which the smaller thumbnail is represented here. The contents of the archive are entirely searchable by author, tags, and almost any conceivable search, such as date, etc. The system provides a clear APA citation for the material, which is useful for future users.
Users are a perpetual question with the archive, while there are interesting academic politics surrounding the archive, its contents, and who can and should use them for what purposes, the real issue is less those politics than the lack of use and users in general. From the initial set of content creators and contributors as indicated in

earlier discussion, now we are basically only receiving spam, which we filter. The material growth of the archive is minimal in the last year, and without continued effort and as such funding, we doubt it will have any more substantial growth. The competition amongst various parties about who does what with what parts of which archive, especially the tensions between research, memorial and archival missions highlight the differences of users of the current and future archive [2]

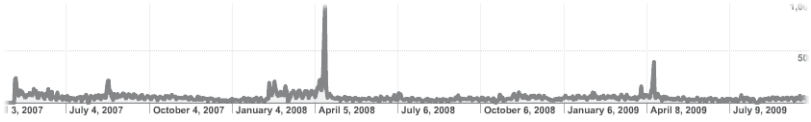
If we think of three sets of users of the archive being contributors/memorialists, researchers including news reporters, and archivists, there seems to be tensions between what they want and what the archive provides. The archive only contains the material contributed and other than two objects, it shows all of that contributed material to the public at any given time. However, the availability of the material does not make it useful to the university researcher who wants to use the material as representing individuals instead of as documentary material, as we did not request rights to use the material to research the creators from the creators. However, over time, the distance between creators and material is fading, much like the memory is fading as described above, and with that fade, the human subjects issues are also fading.

Fading memories and forgetting is quite normal in regard to such tragic events. Forgetting as a social and political process is important for the reconstitution of subjectivities and social, political relations [3][2].

FADING RELATIONS, FADING MEMORIES, FADING INTERESTS

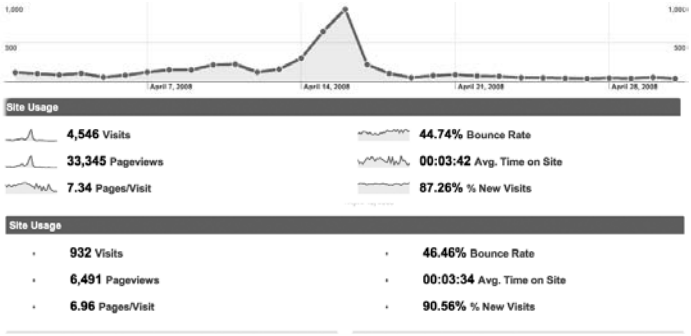
The question that drives this paper and the descriptions so far has been, "what is this archive?", but the question that this paper answers is, "What happened to the users of this archive?". To answer that question, we will inspect and describe the phenomena of our users as represented in the Google Analytics (TM). We use Google Analytics (TM) for user data because they strip out 99% of the bot and otherwise inhuman access, representing the human, and thus representing the marketable to Google, access to the page. By leveraging this tool, I was able to develop reports at several levels of analysis comparing the 2007, 2008, and 2009 data sets available.

The first data object is on the one hand the most revelatory, and on the other hand, the least data intensive to understand. From the date we turned on Google Analytics (TM) for this in May 2007, the data is fairly consistent, with a few peaks, such as the start of school in Fall 2007, the anniversary of the event, the Tragedy of Northern Illinois University on Feb. 14 2008 and the second anniversary of Northern Illinois and third anniversary of the Virginia Tech Tragedy. The peaks indicate high points of use. Correlating with these reference peaks there are contacts from media and other researchers about some topic via email and phone. We can see that other than on such peak events, the number of visits to the archive is relatively low, averaging below 100 on any given day for the past few years, though it has been decreasing over time, which is clear from the data representations.

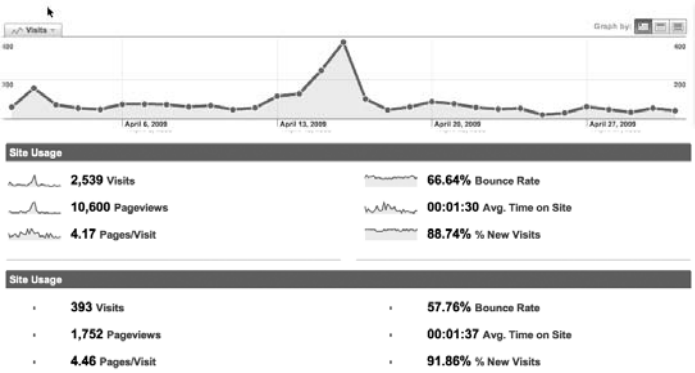


May 3rd 2007 to August 2009 user visits report from Google Analytics(TM)

Investigating the peaks reveals a bit more below, I have representations of two peaks. The peaks represent the second and third anniversaries of the April16archive. These two peaks provide insights particular to the archive's significant event. In the 2008, the first anniversary, we can see fairly interesting behavior for an internet archive, the use of the archive for the month of April was 4546 visits with around 3 minutes and 42 seconds per visits with 33,345 page views and only a 44.74% bounce rate. This means that people are coming to the archive and looking around for a significant period of time, clicking from page to page, from object to object. Surprisingly, the time spent on the site and number of pages per visit goes down for the 932 visits to the site and 6491 page views on April 16, 2008. This is likely related to the increase in new visitors over the average to the month of April. Many people will still remembering and revisiting their content in this period. People unfamiliar with the material in the archive are less likely to spend time on this archive when other archives with different content are also available.



While the low numbers in 2008 are interesting to some extent, they gain their strength as indicators of a loss of contributors/interested parties in relation to the even lower numbers in 2009.



The third anniversary of the April 16 Tragedy, shows the archive getting significantly less traffic in the month of April, at slightly more than 1/2 the visits and 1/3 the page views, the bounce rate is 1/3 higher, and we have fewer pages per visit. Before the third anniversary we did launch the frontpages archive, but that has had little effect to the primarily site. The day of April 16 is significantly less in all statistics in comparison to the prior archive with our key indicators of pages per visit and average time on site, which we take to be indicators of interest to the site and the material falling off significantly.

CONCLUSION

With people spending less time on the site, viewing fewer pages on the site, I feel fairly safe in saying that there is likely less interest in the event across various user groups, and with this loss of interest we have a loss of activity. This loss of activity does not hurt the legitimacy of the archive for researchers, but does likely relate to fading memories and the peripheralization of the event to people's lives. This change seems lessens the interaction with the content creators that contributed materials to the archive, which in terms removes over time some of the web 2.0 orientation of the archive. I suspect that most event-driven archives face the same issues of community and fading memories.

REFERENCES:

[1] Jesiek, B. K., & Hunsinger, J. (2008). THE APRIL 16 ARCHIVE: Collecting and Preserving Memories of the Virginia Tech Tragedy. In B. Agger & T. W. Luke (Eds.), There is a Gunman on Campus (p. 22). Lanham, Maryland: Rowman & Littlefield.

[2] Jesiek, B. K., & Hunsinger, J. (2009). Collecting and Preserving Memories From the Virginia Tech Tragedy: Realizing a Web Archive. In N. Brügger (Ed.), Web Histories. London: Peter Lang.

[3] Auge, M. (2004). Oblivion. University of Minnesota Press.

Abstract
Web content plays an increasingly important role in the knowledge-based society, and the preservation and long-term accessibility of Web history has high value (e.g., for scholarly studies, market analyses, intellectual property disputes, etc.). There is strongly growing interest in its preservation by libraries and archival organizations as well as emerging industrial services. Web content characteristics (high dynamics, volatility, contributor and format variety) make adequate Web archiving a challenge.
LiWA will look beyond the pure "freezing" of Web content snapshots for a long time, transforming pure snapshot storage into a "Living" Web Archive. In order to create Living Web Archives, the LiWA project will address R&D challenges in the three areas: Archive Fidelity, Archive coherence and Archive interpretability. The results of the project will be demonstrated within two application scenarios namely "Streaming Archive" and "Social Web Archive". The Streaming Archive application will showcase the building of an audio-visual Web archive and how audio and video broadcast related web information can be preserved. The Social Web application will demonstrate how web archives can capture the dynamics and the different types of user interaction of the social web.
Keywords: Web Archiving, Rich Media, Spam Detection, Crawl Coherence, Terminology Evolution

Introduction

The Web today plays a crucial role in our information society: it provides information and services for seemingly all domains, it reflects all types of events, opinions, and developments within society, science, politics, environment, business, etc. Due to the central role the World Wide Web plays in today's life, its continuous growth, and its change rate, adequate Web archiving has become a cultural necessity in preserving knowledge. Consequently a strong growing interest in Web archiving library and archival organizations as well as emerging industrial services can be observed.
However, web preservation is a very challenging task. In addition to the "usual" challenges of digital preservation (media decay, technological obsolescence, authenticity and integrity issues, etc.), web preservation has its own unique difficulties:

- distribution and temporal properties of online content, with unpredictable aspects such as transient unavailability,
- rapidly evolving publishing and encoding technologies, which challenge the ability to capture web content in an authentic and meaningful way that guarantees long-term preservation and interpretability,
- the huge number of actors (organizations and individuals) contributing to the web, and the wide variety of needs that web content preservation will have to serve.

A first generation of Web archiving technology has been built by pioneers in the domain like the Royal Library of Sweden and the Internet Archive based on existing search technology. It is now time to develop the next generation of Web archiving technology, which is able to create high-quality Web archives overcoming the limitations of the previous generation. The aim of the European funded project LiWA is to create innovative methods and services for Web content capture, preservation, analysis and enrichment.
In the following section we first give an overview about the current state in Web archiving. Afterwards we will introduce in more detail the Living Web Archives project followed by an overview of the approaches to address the previously mentioned issues. Furthermore we will give an overview of the applications to be developed within the project. Finally the paper concludes and gives an outlook on the remaining project life time.

The Living Web Archives Project

The LiWA project, started in February 2008, brings together a consortium of highly qualified researchers (L3S Research Center, Max Planck Society, Hungary Academy of Science), archiving organizations (European Archive Foun-

Chiara Cirinnà and Maurizio Lunghi (edited by), *Cultural Heritage on line: Empowering users: an active role for user communities. Florence, 15th - 16th December 2009* ISBN 978-88-6453-184-7 (print) ISBN 978-88-6453-187-8 (online)

dation, Sound and Vision Foundation (NL), National Library of the Czech Republic, Moravian Library) and a commercial company (Hanzo Archives). It is the intention of the project partners to turn Web archives from pure Web page storages into "living Web archives" within the next three years. Such living archives, will be capable of: handling a variety of content types; dealing with evolution as well as improving long-term content usability. In order to create Living Web Archives, the LiWA project addresses R&D challenges in the three areas: Archive Fidelity, Archive coherence and Archive interpretability:
Archive Fidelity: development of effective approaches and methods for capturing all types of Web content including the Hidden and Social Web content, for detecting capturing traps as well as for filtering out Web spam and other types of noise in the Web capturing process.
Archive Coherence: development of methods for dealing with issues of temporal Web archive construction, for identifying, analysing and repairing temporal gaps as well as methods for enabling consistent Web archive federation;
Archive Interpretability: development of methods for ensuring the accessibility, and long-term usability of Web archives, especially taking into account evolution in terminology and conceptualization of a domain;
The results of the project will be demonstrated within two application scenarios namely "Streaming Archive" and "Social Web Archive".
LiWA Approaches
In the following sub-section we give an overview about the selected approaches in the four research areas covering the three objectives of the LiWA project. These approaches were developed after getting a detailed understanding of the requirements and the system architecture. The requirements analysis collected the requirements from three different angles. The user angle describes the desirable usage of web archives by libraries and archives. The technical angle collects functional requirements necessary to meet the user requirements of libraries and archives and the intention to extend the current state-of-the-art in/of web archiving. Finally the architecture angle defines functional requirements necessary to integrate LiWA services into one advanced web archiving infrastructure.

Capture of Rich and Complex Web Content
The aim of this working area is to improve dramatically the fidelity of Web archives by enabling capture of content defeating current Web capture tools. This comprises the ability to find links to resources regardless of the encoding using virtual browsing, the detection and capture of structural hidden Web and the capacity to handle streaming protocols to capture rich media Web sites. In order to develop an interpretation/execution-based link extractor for complex and dynamic objects, potential Javascript rendering engines for tasks were identified and tested. The comparison lead to select "WebKit" for implementation as it offers a huge number of features like JavaScript getters and setters, DOM class prototypes, significant JavaScript speed improvements, support of new CSS3 properties. DOM manipulation issues were analysed in depth to develop better links extraction. Various strategies to manipulate DOM from Webkit were tested. The result is a customized version of WebKit for the special use of link extraction.
For capturing rich media open source modules and helper application to support AV applications were tested. The Mplayer was selected as the basis for the helper tool implementation. In order to develop an improved rich media capture module, the crawlers were de-coupled from the identification and retrieval of streams and then moved to a distributed architecture where crawlers communicated with stream harvesters through messages.

Data Cleansing and Noise Filtering
The ability to identify and prevent spam is a top priority issue for the search engine industry [1] but less studied by Web archivists. The apparent lack of a widespread dissemination of Web spam filtering methods in the archival community is surprising in view of the fact that, under different measurement and estimates, roughly 10% of the Web sites and 20% of the individual HTML pages constitute spam.
Spam filtering is essential in Web archives even if we acknowledge the difficulty of defining the boundary between Web spam and honest search engine optimization. Archives may have to tolerate more spam compared to search engines in order not to loose some content. Also they might want to have some representative spam either to preserve an accurate image of the Web or to provide a spam corpus for researchers. Therefore the main objective of spam cleansing in Web archives is to reduce the amount of fake content the archive will have to deal with. The envisioned toolkit will

help prioritize crawls by automatically detecting content of value and exclude artificially generated manipulative and useless content.
The current LiWA solution is based on the lessons learned from the Web Spam Challenges [2]. As it has turned out, the feature set described in [3] and the bag of words representation of the site content [4] give a very strong baseline. Therefore the LiWA baseline content feature set consists of the following language-independent measures: the number of pages in the host, the number of characters in the host name, in the text, title, anchor text etc; the fraction of code vs. text, the compression rate and entropy; and the rank of a page for popular queries. Within this set we use the measures for in- and outdegree, reciprocity, assortivity, (truncated) PageRank, Trustrank [5] and neighborhood sizes, together with the logarithm and other derivatives for most values. Whenever a feature refers to a page instead of the host, we select the home page as well as the maximum PageRank page of the host in addition to host-level averages and standard deviation.
In addition LiWA services intend to provide collaboration tools to share known spam hosts and features across participating archival institutions. A common interface to a central knowledge base will be built in which archive operators may label sites or pages as spam based on own experience or suggested by the spam classifier applied to the local archives.

As a major step in disseminating the special needs of Internet Archives, we propose tasks for a future Web Spam Challenge [6]. We generate new features by considering the temporal change of several crawl snapshots of the same domain [7]. In addition by the needs of collaboration across different archival institutions we also provide training labels over one top level domain and request prediction over a different domain.

Archive Coherence
A common notion of "coherence" refers to the explanations given in the Oxford English Dictionary (cf. http://dictionary.oed.com) describing coherence as "the action or fact of cleaving or sticking together", which - in terms of a Web site - results in a "harmonious connexion of the several parts, so that the whole 'hangs together'". From an archiving point of view, the ideal case to ensure highest possible data quality of an archive would be to "freeze" the complete contents of an entire Web site during the time span of capturing the site. Of course, this is illusion and practically infeasible. Consequently, one may never be sure if the contents collected so far are still consistent with those contents to be crawled next. However, temporal coherence in Web archiving is a key issue in order to capture digital contents in a reproducible and, thus, later on interpretable manner. To this end, we are developing strategies that help to overcome (or at least identify) the temporal diffusion of Web crawls that last from only a few hours up to several days. Therefore, we have developed a coherence framework that is capable of dealing with correctly as well as incorrectly dated contents[8]. Depending on the data quality provided by the Web server, we have developed different coherence optimizing crawling strategies, which outperform existing approaches and have been tested under real life conditions. Even more, due to the development of a smart revisit strategy for crawlers we are also capable of discovering and (as a consequence) of ensuring coherence for contents, which are incorrectly dated and thus not interpretable with conventional archiving technologies. Current results make temporal coherence of Web archiving traceable under real life applications and provides strategies to improve the quality of Web Archives, regardless of how unreliable Web servers are.

Archive Interpretability
The correspondence between the terminology used for querying and the one used in content objects to be retrieved is a crucial prerequisite for effective retrieval technology. However, as terminology is evolving over time, a growing gap opens between older documents in (long-term) archives and the active language used for querying such archives. Language changes are triggered by various factors including new insights, political and cultural trends, new legal requirements, high-impact events, etc.
An abstract model has been developed [9] that allows the representation of terminology snapshots at different times (term-concept-graphs). From this we derived that the act of automatically detecting terminology evolution given a corpus can be divided into two subtasks. The first one is to automatically determine, from a large digital corpus, the senses of terms. Such a word sense discrimination module has been implemented and successfully been tested on

the Times corpus that covers 200 years of news articles. Current work focuses on the second step – the detection of terminology evolution. In this step the word clusters detected in the first step are tracked over time to detect evolution and to derive mappings.

Applications

The LIWA Technologies can be used either at crawl-time or after completion of the crawl, integrated with existing web archiving workflow. In order to test and apply these new methods and results, an integration platform of the modules is being built both by the European Archive Foundation (using open source tools) and by Hanzo Archives.
Two applications scenarios are developed in LIWA to illustrate the possible use of these technologies in real world scenario whose scope is wider than what LiWA specifically addresses.

LiWA technology for content and context in Sound and Vision archive

The Netherlands Institute for Sound and Vision is one of the largest audio-visual archives in Europe. The cultural heritage preservation policy of the Institute implies that the AV archive should preserve the Dutch audiovisual cultural heritage. As the Internet is increasingly becoming an important source for (user generated) audiovisual cultural heritage content, Sound and Vision has a strong commitment to capture information available on the Web. More specifically, the institute is eager to capture broadcast related websites, including streaming content. However, as capturing streaming content from the web is difficult, until now only a selection of user generated video content is downloaded manually from the Internet. With the streaming content capturing technology developed in the LiWA project, Sound and Vision is able to address the capturing of Dutch cultural heritage content in a much more efficient way.
Besides being a potential provider of audiovisual content, the Web is regarded as a valuable source for gathering contextual information that relates to the collections. Context information is relevant for both documentalists, and also other users interested in a specific broadcast or a broadcasting related topic, such as journalists, teachers or researchers. Typically, these users have to use different interfaces for different sources to search these sources. Ideally, Sound and Vision provides these users with a single interface that allows searching both the digital asset management system of the AV archive (iMMix) and related web content. The LiWA application Streaming demonstrates how broadcast related potential end users could access web content. The archived content will be used as test data for the development of the Sound and Vision context data platform that specifically addresses the linking of web context to the digital asset management system of Sound and Vision.

Social web application

Social web sites typically contain highly inter-linked content and use dynamic linking, widgets and tools as well as high degree of personalisation. Capturing social web sites is extremely challenging and cannot be fully achieved using current methods and tools. Social web thus represents one of the greatest challenges in web archiving.
With the Social web application, LiWA intends to demonstrate a dramatic improvement in both archive structure and content completeness so that the rapidly evolving and increasingly diverse content of the social Web is captured more accurately and evenly. The aim of the application is to show how the LiWA technology fits in the workflow of an active Web archiving institution, by considering a real-life scenario of the National Library of the Czech Republic. The application is designed as a set of independent modules developed in LiWA as described in section 2. The modules can be readily integrated with existing Web archiving workflow management tools. A Web archiving institution can choose to deploy all of the modules or just some of them, depending on its needs and particular workflow. The application is designed as generic and can be used to enhance archiving of any type of web content, not just social web.

Conclusions & Outlook

In this paper we presented important issues in Web archiving and introduced the Living Web Archives project, which aim is to overcome these limitations. For research areas have been identified namely Capturing of Rich and Complex Web content, Data Cleansing and Noise Filtering, Archive Coherence and Archive Interpretability. Promising solutions have already been developed and continuously being enhanced in the second half of the project. Furthermore the presented application showcases will be implemented.

ACKNOWLEDGMENTS

This work is funded by the European Commission under LiWA (IST FP7 216267).

REFERENCES

[1] M. R. Henzinger, R. Motwani, and C. Silverstein. Challenges in web search engines. SIGIR Forum, 36(2):11–22, 2002.

[2] C. Castillo, K. Chellapilla, and L. Denoyer. Web spam challenge 2008. In Proceedings of the 4th International Workshop on Adversarial Information Retrieval on the Web (AIRWeb), 2008.

[3] C. Castillo, D. Donato, A. Gionis, V. Murdock, and F. Silvestri. Know your neighbors: web spam detection using the web topology. Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval, pp. 423–430, 2007.

[4] I. Bíró, D. Siklósi, J. Szabó, and A. A. Benczúr. Linked latent Dirichlet allocation in web spam filtering. In AIRWeb '09: Proc. 5th int. workshop on Adversarial information retrieval on the web, 2009.

[5] Z. Gyöngyi, H. Garcia-Molina, and J. Pedersen. Combating web spam with TrustRank. In Proc. of the 30th Int. Conference on Very Large Data Bases (VLDB), pp. 576–587, Toronto, Canada, 2004.

[6] A. A. Benczúr, M. Erdélyi, J. Masanés, and D. Siklósi. Web spam challenge proposal for filtering in archives. In AIRWeb '09: Proc. of the 5th Int. Workshop on Adversarial Information Retrieval on the Web. ACM Press, 2009.

[7] M. Erdélyi, A. A. Benczúr, J. Masanés, and D. Siklósi. Web spam filtering in internet archives. In Proc. of 5th the Int. Workshop on Adversarial information retrieval on the web (AIRWeb), 2009.

[8] M. Spaniol, D. Denev, A. Mazeika, P. Senellart and G. Weikum. Data Quality in Web Archiving. In Proceedings of the 3rd Workshop on Information Credibility on the Web (WICOW 2009) in conjunction with the 18th World Wide Web Conference (WWW2009), Madrid, Spain, April 20, 2009, pp. 19-26.

[9] N. Tahmasebi, T. Iofciu, T. Risse, C. Niederee, and W. Siberski; Terminology Evolution in Web Archiving: Open Issues; In Proc. of the 8th Int. Web Archiving Workshop 2008, Aarhus, Denmark.

Abstract.
In Belgium, many institutions have a lot of information stored on analogue carriers. This information is likely to get lost if no digitized copy of the information is stored for the long term. Long-term preservation is subjected to many risks. Overcoming those risks starts with describing the data thoroughly. The metadata needed for long-term preservation are descriptive metadata to search and manage the whole archive, binary metadata to describe the bitstreams, technical metadata describing the files, structural metadata for the representation information, preservation metadata for keeping track of the provenance of the data, and rights metadata. Therefore, we developed a layered semantic metadata schema. The top layer holds the descriptive metadata, the bottom layer holds all the information necessary for long-term preservation. The top layer consist of an OWL representation of Dublin Core, while for the bottom layer we developed an OWL representation of the preservation standard PREMIS 2.0, extended with a vocabulary defining the legal roles of a person, organization, or software. This way, our model offers all the necessary metadata for long-term preservation.
Keywords: digital preservation, PREMIS 2.0, ontology, semantic web

Introduction

In Belgium, the broadcasters, cultural organizations, private persons, and government institutions possess thousands of hours of speech and image material which is stored on analogue carriers. This material belongs to the most important cultural heritage in Flanders. At this moment, the analogue carriers are degrading and are continuously losing quality, making the data inaccessible. Disseminating and storing the content digitally overcomes this problem only temporarily. Furthermore, this digital content has to remain intact and accessible over time, e.g., 20, 50 years or longer. Digital long-term preservation forms the solution for this issue. The project BOM-VI (Preservation and Disclosure of Multimedia Data in Flanders, [1]) initiates the digital long-term preservation of the cultural heritage in Flanders and researches the problems encountered with digital long-term preservation.
In this paper, we present our layered semantic metadata model. First, in chapter two, we introduce the different kinds of metadata that are needed to overcome all the risks involved in long-term preservation and show how our proposed, layered, semantic metadata model relates to those risks. The semantic model consists of two layers: the top layer delivers the descriptive metadata, and the bottom layer is responsible for the binary metadata, the technical metadata, the structural metadata, the preservation metadata (provenance metadata, fixity metadata, and context metadata), and the rights metadata. This way, all the metadata for describing the content for the long-term, are covered by the layered semantic metadata model. For the top layer, we use a Web Ontology Language (OWL, [2]) representation of Dublin Core [3], which is described in chapter three. For the bottom layer, depicted in chapter four, we developed an OWL representation of the preservation standard Preservation Metadata, Implementation Strategies 2.0 (PREMIS 2.0, [4]). This PREMIS OWL schema (PREMIS OWL, [5]) not only covers the necessary metadata described in chapter two, but also stores the semantics of the metadata for the long term. This can be very important due to, e.g., terminology changes. This schema is accompanied by a vocabulary describing the legal roles that a person, organization, or software application can have.

Metadata levels for long-term preservation

When preserving digital multimedia data for the long term, the digital archive demands some specific requirements. On the one hand the software and hardware of the digital archive have to guarantee access to the information during a long time. On the other hand human input is necessary in the form of archive descriptions, work processes, and the use of standards to keep the information accessible and interpretable as long as possible to the user community. Based

Chiara Cirinnà and Maurizio Lunghi (edited by), *Cultural Heritage on line: Empowering users: an active role for user communities. Florence, 15th - 16th December 2009* ISBN 978-88-6453-184-7 (print) ISBN 978-88-6453-187-8 (online)

on the Open Archival Information System (OAIS, [6]) reference model, the data has to be described on three levels to guarantee long-term preservation. On every level, there are possible risks for loss of data, which can be minimized by describing the data thoroughly.
On the lowest level, a digital file consists of bits and bytes which can change by external influences, like corruption of carriers, migrations, etc. On this lowest level, binary metadata and fixity metadata are needed to correct these errors and to guarantee authenticity of the data.
On a higher level, file formats and compression formats like AVI, MP3, and JPEG describe the way the bits can be transformed to an interpretable multimedia representation. When a file format becomes obsolete, the archive has two solutions to preserve the stored data: migration or emulation. Metadata is needed to support these actions. At this level, it is also very important to preserve the look and feel of the objects. when migrating file formats. Thus, a rich description of the look and feel is also necessary. For this level we need technical metadata, for describing the files, structural metadata, for describing sets of files and their relations, e.g., a book which is represented as a set of scanned TIFF images, and provenance metadata, for describing the history of the content information: the original owners of the data, the processes that determined the current form of the data, and the available versions.
On the highest level, the information should remain interpretable. Institution structures, terminologies, the designated community, and the rights of an object or institution can change over time. To keep the information interpretable, enough information should be included in the archived package. At this level, the archive needs descriptive metadata, for a general description of the object, e.g., MARC, rights metadata, for describing copyright statements, licenses, and possible grants that are given, and context metadata, for describing the relations of the content information to information which is not packed in the information package. Examples of context metadata are related datasets, references to documents in the original environment at the moment of publication, helper files, and the language.
When developing a metadata schema for the long-term preservation of digital multimedia, metadata descriptions on all levels have to be taken into account, going from bit level descriptions to descriptions of the intellectual content. To realize this, we developed a layered semantic metadata schema. The top layer offers the descriptive metadata. The bottom layer takes care of the preservation metadata, rights metadata, binary metadata, technical metadata, and structural metadata necessary for deep archiving. For the top layer, an OWL representation of Dublin Core is developed. For the bottom layer, an OWL representation of the preservation standard PREMIS 2.0 is developed. This standard is based on the OAIS reference model. This schema describes the data on all necessary levels.

Top layer: Dublin Core

Descriptive metadata describes the content of the data: subject, author, date of creation, file format, etc. This metadata makes it possible to manage and search the complete digital archive. When archiving data coming from different sectors like the broadcast sector, the libraries, the cultural sector, and the archival sector, a problem arises concerning descriptive metadata. Many of the institutions already have descriptive metadata. Are these descriptive metadata stored as metadata or as data? Both strategies have their advantages and disadvantages. When archiving these descriptions as metadata, the archive has to provide a metadata schema. The choice of this schema is a non-trivial task. The metadata schemes used for the descriptions are very domain-specific. To store the descriptive metadata lossless the descriptive metadata schema should be some kind of smallest common multiple of all the descriptive metadata schemes offered by the institutions. This would be a huge metadata schema, impossible to maintain. That is why the descriptive metadata is archived along with the data in their original metadata format, e.g., MARC , so there is no information loss. On top of this metadata, the archive offers a broadly accepted descriptive metadata schema. This gives the archive the necessary tools to search the whole archive. When finding the data of interest, the original metadata that is stored as data can still be presented to the users.
Dublin Core was chosen to describe this top layer of descriptive metadata. Dublin Core is a broadly accepted descriptive schema. The power of this schema is its simplicity and generality. It consists of fifteen fields among which creator, subject, coverage, description, date. It can answer to the basic questions: Who, What, Where, and When. All the fields in Dublin Core are optional and repeatable. This makes it possible to map relatively easily almost all the descriptive metadata schemes to Dublin Core whereas many institutions already support Dublin Core.

Bottom Layer: PREMIS OWL

For this layer, we developed an OWL schema of the preservation standard PREMIS 2.0. PREMIS is a preservation standard based on the OAIS reference model. The preservation standard is described by a data model. The data model of PREMIS consists of five semantic units or classes important for digital preservation purposes:

- Intellectual Entities: a part of the content that can be considered as an intellectual unit for the management and the description of the content. This can be for example a book, a photo, or a database.
- Object: a discrete unit of information in digital form.
- Event: An action that has an impact on an object or agent.
- Agent: a person, institution, or software application that is related to an event of an object or is associated to the rights of an object.
- Rights: description of one or more rights, permissions of an object or agent.

Intellectual entities, events, and rights are directly related to an object. An agent can only be related to an object through an event or through rights. This way, not only the changes to an object are stored, the event involved in this change is also described. These relationships offer the necessary tools to store the provenance of an object properly. Fig. 1 clarifies the data model of PREMIS.

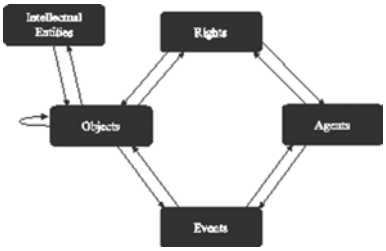


Fig. 1. Data model of PREMIS

Object

The Object class describes a unit of information in digital form. It is related to the intellectual entity class. This intellectual entity is described by descriptive metadata. This descriptive metadata are very domain-specific. For this, there exist already a lot of descriptive metadata models. Therefore, the description of the intellectual entity is out of scope for PREMIS. In our implementation, the top layer describes the intellectual entity.

An Object class knows three subclasses:

- File: a file is an ordered sequence of bytes that is known by the system.
- Bitstream: a bitstream is the actual data inside a file.
- Representation: a representation is a set of files with structural metadata needed for a complete description of an intellectual entity.

The Object class possesses all the necessary features to describe the object on the different levels. The minimum information for describing an object (File, Bitstream, or Representation) are objectIdentifier, which gives the identifier of the object, objectCharacteristics, needed for the Bitstream subclass and the File subclass, which gives the necessary technical and binary metadata, and storage, necessary for describing a File or Bitstream, which indicates either the location the data is stored, either the medium the data is stored on. An object can be decribed further into detail using preservationLevel, because some repositories offer the opportunity to define a preservation level for an object, significantProperties, defining some significant properties of the object, which need to be preserved when, e.g., migrating the data, originalName, for indicating the original names of the packages delivered to the repository, environment, which describes the environment the user needs to render the content and interact with the content, signatureInformation, for

storing digital signatures generated during ingest into the repository, and finally, relationship, which relates to structural metadata to assemble complex objects.
For linking object information to events, intellectual entities, or rights statements, the object class offers three properties, i.e., linkingEvent, linkingIntellectualEntity, and linkingRightsStatement.

Event
An event aggregates all the information about an action that involves one or more objects. This metadata is stored separately from the object metadata. Actions that modify objects should always be recorded as events.
The Event class is described at least by an eventIdentfier, eventType, e.g. capture, creation, and an eventDateTime. This information can be extended using the eventDetail property, which gives a more detailed descrciption of the event, and the eventOutcomeInformation, which describes the outcome of the event, in terms of success, failure, or partial success. These properties are able to describe any event altering an object. The Event class can be related to an Agent class or Object class via the resp. properties linkingAgent and linkingObject.

Agent
This class aggregates information about attributes or characteristics of agents. Agents can be persons, organizations or software. This class provides the necessary tools to identify unambiguously an agent. The minimum properties needed to describe the Agent class are agentIdentfier and agentType. Optionally, an agent can also be described using the agentName. This is just enough to identify the agent.
An agent can hold or grant one or more rights. It may carry out, authorize, or compel one or more events. An agent can only create or alter an object through an event or with respect to a rights statement. The relationships between an agent and an object through an event or rights entity make it possible to describe the whole provenance of an object.

Rights
The minimum core rights information that a preservation repository must know, is what rights or permissions a repository has to carry out related to objects within the repository. These may be granted by copyright law, by statute, or by a license agreement with the rights holder. Rights entities can be related to one or more objects and one or more agents. Every Rights class can be related to different RightsStatements. A RightsStatement knows three subclasses: the Copyright subclass, the License subclass, and the Statute subclass. These three subclasses offer the necessary metadata for describing, rights information, i.e., copyrights, licenses, and statutes. Every RightsStatement is decribed at least by a rightsStatementIdentifier, and has also the optional property rightsGranted, which describes the actions the granting agency has allowed the repository. The RightsStatement class can be related to an Object class or Agent class via the optional, repeatable object properties: linkingObject and linkingAgent.
This part of the PREMIS OWL schema is extended with a vocabulary that describes the roles agents can have concerning a rights statement. This vocabulary is based on the results of research performed within the project BOM-VI. To fully describe the rights of an object, all the persons, involved in the production of the described object, should be taken into account which is for many organizations impossible. Therefore, a checklist was made with the most important rights and rights holders that should be described. Based on this checklist a vocabulary was made to describe these important legal roles of an agent, e.g., author, composer, conductor.

CONCLUSION
When preserving digital information for the long term, different metadata are important. Descriptive metadata are needed to describe the intellectual entities, binary metadata, technical metadata, and structural metadata are essential for the description of the data on all levels (bitstream, file, representation). Preservation metadata is necessary to describe the provenance of the data, to guarantee the authenticity of the digital data, and to provide a context. At last, rights metadata also needs to be stored.
The two-layered, semantic metadata schema described in this paper offers all these metadata. The top layer takes care of the descriptive metadata. An OWL representation of DC was chosen for this layer. The bottom layer carries the binary metadata, technical metadata, structural metadata, preservation metadata, and the rights metadata. For this

layer an OWL representation of PREMIS 2.0 was developed. To describe the rights in a more detailed manner, the PREMIS OWL schema was extended with a vocabulary defining the different legal roles of persons, organizations and software. By describing the data with this layered metadata schema, all the risks that come with long-term preservation are minimized. By splitting up the semantic schema in two layers, the top layer with the descriptive metadata can be made public and weaved into the web of data, if the rights permit it. The bottom layer remains closed for the public and is responsible for the long-term preservation of the data.

References

[1] Preservation and Disclosure of Multimedia Data in Flanders, https://projects.ibbt.be/bom-vl/
[2] Dean, M., Connolly, D., van Harmelen, F., Hendler, J., Horrocks, I., McGuinness, D. L., Patel-Schneider, P. F., Stein, L. A.: OWL web ontology language reference, W3C Working Draft, http://www.w3.org/TR/2003/WD-owl-ref-20030331 (2003)
[3] The Dublin Core Metadata Initiative, DCMI, http://dublincore.org/ (2009)
[4] Higgins, S.: PREMIS Data Dictionary, Digital Curation Centre (DCC), http://www.loc.gov/standards/premis/v2/premis-dd-2-0.pdf, Glasgow (2007)
[5] Coppens S., Mannens E., Van de Walle R.: PREMIS OWL, Semantic Model of PREMIS 2.0, http://multimedialab.elis.ugent.be/users/samcoppe/ontologies/Premis/premis.owl
[6] Consultative Committee for Space Data Systems (CCSDS): Reference Model for an Open Archival Information System. Blue book. Issue 1, 148 p., CCSDS, Washington (2002)

FP7 Integrated Project, third call
Started on January 2009, duration 40 months
Partners : INA (F), BBC (GB), RAI (It), JRS (Au), B&G (Nl), ORF (A), ExLibris (Israél), Eurix (It), Doremi (F), Technicolor (Nl), IT Innov. (GB), Vrije Universiteit Amsterdam (Nl), Universität Innsbruck (Au, European Digital Library Foundation (Nl)

Getting into the digital world is a highly challenging issue for any Audiovisual collection due the complexities and costs of digital transfer. However a still more challenging issue is how to remain in the digital world within its continually changing context. Information systems, formats, definitions change continuously and induce regular actions on contents, descriptions and systems so to guarantee access on the long-term. There is no clear vision how all these actions affect contents or how to ensure that these actions will not modify them.
For Digital Libraries this represents a major evolution in their conception: initially build around digitised books, they have started including all kinds of cultural digital contents from Archives, Museums and Audiovisual Libraries. The main challenge in past years was bringing analogue contents to the digital world, then, conceiving specific tools for the management, description and structuring of digital contents. The challenge today for content holders is to make all contents available to the citizen, in secured environments and respecting intellectual property. Major European projects advance in this direction through intelligent access portals with millions of contents.
A large array of problems still remains within digital libraries; they may be particular to kind of content or media or transversal to different domains. They concern the permanence of Digital objects through time (a major concern for any digital content owner); Interoperability of contents and metadata; management and handling of Rights; Content tracking and identification and; a less technical however important issue: how to economically foster major digitization actions and digital preservation.

Specific domain

The Audiovisual domain presents challenges and issues that need a specific approach:

- The mass of accumulated material since the beginning of film and broadcast industry is huge (estimated in 100 million hours)
- Most of it is still in an analogue format
- It is totally dependent of access technology (machines, readers)
- Digital born material presents similar problems of technological access and format dependency
- Professional audiovisual material increases, in Europe, at a rate of circa 5 million hours per year
- Audiovisual contents need efficient and extensive metadata for archival and access purposes
- They represent huge storage volumes (1 million hours of video at a 1mb/s rate, which is a non professional rate, represent 3,6 Petabytes of storage)
- It generally presents complex right ownership situations or poorly identified ownership
- If not managed with a long-term perspective they are easily subject to loss or inaccessibility
- Audiovisual archiving represents a continuous cost for content holders, archiving needs to be carefully evaluated, planned and implemented

Past projects like Presto and PrestoSpace, have developed specific technologies and business plans in order to address, on an industrial basis, the problem of analogue to digital migration. The results of the project have been brought back to the community through preservation machines, tools for storage management, tools for restoration, metadata platforms, business models and strong methodologies for an industrial approach .

Chiara Cirinnà and Maurizio Lunghi (edited by), *Cultural Heritage on line: Empowering users: an active role for user communities. Florence, 15th - 16th December 2009* ISBN 978-88-6453-184-7 (print) ISBN 978-88-6453-187-8 (online)

PrestoSpace opened the gates to analogue to digital migration through its achievements and the strong price-reducing factor it brought to the activity in the domain. It also structured the actors of the community and set the basis of an efficient interaction among them. However these important steps, as essential as they have been, are not enough to inscribe the preservation of Europe's Audiovisual Heritage on the very long term. Managing and securing huge masses of digital contents with an appropriate description and identification is an indispensable step to assure content owners of the durability of their assets and to promote large access programs to contents.

PrestoPRIME: an Integrated Project for Digital Audiovisual Contents

Getting into the Digital world represented the first indispensable step in order to address the problem of Digital Audiovisual Archives as a whole. A large awareness has been built on the necessary actions that need to be done in order to avoid the analogue black hole. Preservation and documentation functionalities have been improved and brought to common working environments and new semantic tools are being largely developed to improve access, recognition and identification of contents.

Still, mainly due to the volumes, the complexity and the diversity of the audiovisual domain; there is a strong need for research and development of tools and methodologies to assure what is the new challenge for digital contents: long-term preservation, identification and access. The considerable growth of the number of audiovisual contents from the past and regularly produced by professionals and non-professionals, associated to the increasing circulation of those images, introduces two new problems that deal directly with preservation:

The origin and identification of contents: Who made it? Who does it belong to? Can I use it? Where is the original? Can I find it in a better quality?

Content archiving and preservation: Who is keeping them? Does he have the mission to do it? With which time perspective? Will the contents be there in a hundred years? Will I be able to access them? How can I secure my own contents?

The PrestoPRIME project addresses the new challenges that need to be tackled in order to guarantee long-term access and usability of contents. The project is structured in four domains that constitute a global approach to an organised structure for audiovisual contents permanence. It brings together two research domains concerning

- Digital permanence, long-term storage and content identifications
- Interoperability and quality assessment
- These central aspects are dealt in close relationship with two other major issues which are:
- A Right management environment to model European right legislation and propose effective tools for content exchange at a European level
- A Competence Centre dedicated to Audiovisual preservation, restoration and documentation issues, serving as a reference institution for content holders, service providers, industrials and research projects.

The Competence Centre, a central concept for PrestoPRIME

The Competence Centre is a major tool for the PrestoPRIME project, its function is to foster, accelerate and become a reference for Preservation and Migration actions in the Audiovisual domain. Based on the results of previous projects for all what concerns analogue to digital migration; it will incorporate the results of research and development issued from the project, thus being functional from the beginning of the project. Furthermore, its role as a federator of actors within the audiovisual domain: -audiovisual archives and collections - service providers – industrials - academic and industrial research; its structuring actions: - registers of experts – registers of technologies – registers of works; will make the Competence Centre become a central actor for the domain, in relation the major access projects like Europeana and in close relationship with the European Commission.

The Competence Centre will be launched in October 2010 through its portal.

Abstract
Contemporary musical production makes an heavy usage of digital artefacts, either hardware or software based. Since the middle of the 80es, an important issue has been recognized: the fast obsolescence of hardware and software products endanger seriously the future of this production. The question is not only to preserve the results, by recording them, but to preserve the ability to reperform the works live, as we do today for music of the last centuries.
We will present the methodology we develop in the ASTREE project for building knowledge in relation to musical works, and particularly for digital processes that are considered as specific music instruments. We will discuss the different issues in preservation of contemporary music, and show that one of the most prominent is the lack of formalized knowledge about the digital musical instruments, their notation, and their integration into musical score. We will present our efforts towards building an organology of real-time audio processes, and show that this can be the basis for an adequate musical notation and its integration in musical score.
Keywords : Music, Contemporary, Digital, Preservation

Introduction

A brief history

The first interactive works combining performers and realtime electronic modulation of their parts have appeared in the middle of the 80es. Electronic devices, either hardware or software, have been interfering with various musical configurations: the instrument-computer duet, for instance in Manoury's works (Jupiter, for flute and computer, 1987-1992 ; En Echo, for voice and computer, 1993-1994); the works for ensemble and live electronics, such as Fragment de lune (1985-1987) by Philippe Hurel; the works for soloists, ensemble and electronics, such as Répons (1981-1988) by Pierre Boulez.
After nearly 25 years of interactive works, institutions have become aware that this type of music is completely dependant on its hardware and software implementation. May the operating system or the processor evolve, the piece cannot be reperformed. This is for instance what nearly happened to Diadèmes, a work by composer Marc-André Dalbavie for alto solo, ensemble and electronics. First created in 1986 and honoured by the Ars Electronica Prize, the work was last performed in 1992. In december 2008, its american creation was planned, more than 22 years after its premiere in France. But the Yamaha TX 816 FM synthetizers previously used are no longer available, and the one still present at Ircam is nearly out of order. Moreover, composer Dalbavie has tried several software emulators, but none of them was suitable according to him to replace the old hardware synthetizer.
In april 2008, Dalbavie and his musical assistant Serge Lemouton decided to choose another technique: they built a sampler. It is a kind of database of sounds produced by an instrument. The sounds have been recorded from the old TX 816 at various pitches and intensities. This solution enabled to reperform the piece having a kind of photography of the previous sounds. When no sound corresponding to a given pitch exists, the sampler is able to interpolate between existing files, to give the illusion the missing note exists.
One quickly understands the maintenance activity to be able to reperform a work is a never ending activity that should moreover respect a minimum of authenticity.

Aims of preservation

Having in mind that the aim of preservation is to make possible new performances of the works, it becomes clearly not sufficient to preserve the outputs – audio or video recordings – even if these recordings are clearly part of the objects to be preserved.

Chiara Cirinnà and Maurizio Lunghi (edited by), *Cultural Heritage on line: Empowering users: an active role for user communities. Florence, 15th - 16th December 2009* ISBN 978-88-6453-184-7 (print) ISBN 978-88-6453-187-8 (online)

At the very core of performance is the real-time process, often called “the patch” : it is software that takes data in input – directly from the performer, or from prerecorded data, audio or video, and process them before rendering the output on speakers or a display. The real-time process is the expression of the ideas of the composer regarding the use of digital material, it is then the main object to be preserved, is addition to the score (French composer Philippe Manoury often calls the digital material he is using a “Digital Orchestra”).

Risks and strategies

The different strategies

Active preservation of realtime interactive music involves various aspects and is based on various actions. The first step is the physical conservation of all elements necessary for the reperformance of the work: the score, the patches, the various instructions, etc. At Ircam, the Mustica server provides patch files and instructions of implementation for a selection of nearly 60 works.

Another possible strategy is emulation, definitely one of the most difficult. Bernardini and Vidolin [1] quote the example of Stockhausen’s Oktophonie, which requires an ATARI-1040 ST computer that no longer exists. There are Atari emulators running on other computers but nobody knows whether the Notator program used by Stockausen will run on an emulator, though communities of users may give some help…

Migration is the most widespread activity to achieve reperformance. Many composers had their works transformed from one technical environment to another. All institutions in the field of electronic arts face migration necessities. At IRCAM, important pieces using Next computers were moved to Macintosh machines at the end of the 1990s.

Last but not least, virtualization means describing electronic modules using abstractions. At IRCAM, Andrew Gerzso has completed an important work aiming at finding representations as independent as possible from technical implementation for signal processing modules in Anthèmes 2, by Pierre Boulez, for violin and live electronics. The effects used in the piece have been added to the score as if they were instrumental parts. This is according to us the ultimate level of virtualization. It has also to be noticed that current musical notation (Common Western Musical Notation) is virtual, in the sense that it is independent from any implementation : we can play music written for any instrument on another instrument, the paradigms of notation being sufficiently abstract in order to achieve this goal.

The musical notation issue

The need to integrate new technology in the musical score has been recognized a long time ago [2]. But, despite numerous tries, it seems that few systems have emerged. One can for instance examine the problem of notation for spatialization.

Spatialization of sound seems to be now a well-known domain, where numerous realizations have been made, and that offers a wide range of experiments. But there are few certainties, few theoretical studies and few references [3].

Concepts and terms used are vague, not precisely defined, and their acceptions are different according to the point of view of the actor, depending on numerous factors [3] :

- actions have different meanings according to the point of view : producer (an audio engineer...) or receiver (a listener)
- descriptors have different meanings according to the point of view envisioned : reality, image of the reality, or conceptualization of reality.

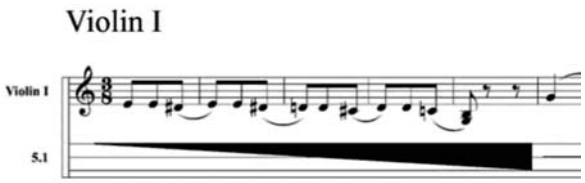


Figure 1 : a notation of spatialization for a 5.1 system

In this context, it becomes very difficult to envision a musical notation that takes into account all these different point of views, before having realized a unification, or merely a standardization of the domain, putting in relation the different points of view expressed.
Moreover, some tries to achieve notation of spatialization of the score becomes dependent on the physical implementation, like the following one that is dependent on a 5.1 system [4]:

The need of a rationalized approach
From the remarks exposed above, we recognized the need to build a common base about the digital musical instruments, from where we could extract and constitute the knowledge basis for a rationalized approach of the musical notation issue. This approach is the basis of the ASTREE project that is exposed below.

THE ASTREE STRATEGY

The ASTREE methodology
The ASTREE methodology is twofold:
First, existing processes have to be translated into a common language. Second, from that language, we will rebuild the original processes, or equivalent one, we will analyze them by applying data mining techniques, and we will also generate an automatic documentation.
The ASTREE methodology can be summarized in the following figure:

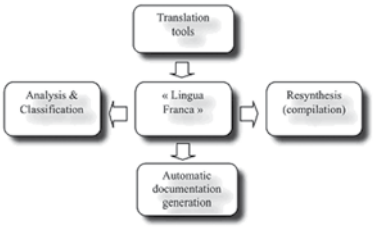


Figure 2 : the ASTREE methodology

Lingua Franca
At the very core of the methodology is a common language, a “Lingua Franca”, that should be completely independent from any hardware or software. Furthermore, this language should be sufficiently expressive to convey the meaning of current existing real-time processes, and at the same time very concise in order to be easily analyzed.
The FAUST language, developed in Grame since 2000 [5], is partially consistent with these requirements. It is a signal processing language, expressed as an algebra, that is aimed originally to process audio signals at a fixed rate, but is currently extended in order to become able to process vectors and matrices, with multi-rate capacities.
The FAUST language is sufficiently expressive for the expression of current objects in the domain of audio processing, at least for the synchronous part.
Translation
We develop tools for translating currently existing processes, built with current environments in use for contemporary music, for instance Max/MSP, but also for other environments like PureData (or even MatLab).
These translation tools are essential, not only for translation of existing material, but also for future material. Users, like composers or computer music designers that assist the composer in his task, are unlikely to use directly an algebraic language like the FAUST language. They will certainly continue using graphical programming environments like Max/MSP or PureData that let them free to experiment and build tools by successive refinements, rather than expressing tools in a language that imposes more or less a preliminary analysis and modeling.
Resynthesis
On top of this language, we can then add the ability to build processes, that in turn will become dependent on the machine, but that can be immediately compared to the original in terms of results: one can compare the output given

by the newly implemented process the original output, as soon as the original process and its translation in the Lingua Franca are available.
The FAUST language can currently be translated in C++, and then compiled in a machine executable. We will also adapt the translation tool in order to work in the reverse order, and obtain Max/MSP or PureData implementations from the FAUST expression.
The purpose of this resynthesis is not only to be used in new performances, but, at the time of archiving processes, to proove that the FAUST expression of the process is sufficient. We can prove it through two stages : first, by doing a reverse translation in the original language, we can prove that no information was lost in the process, and second, by doing a new implementation, and comparing outputs with the original, we can make an a priori evaluation of the authenticity of the translated process towards original.
Automatic documentation generation
The code obtained as well as the source code can be analyzed in order to extract from there an automatic documentation. We can document input and output data, control parameters, extract comments and structures, and generate figures and mathematical expressions corresponding to the source.
Analysis and classification
By applying data mining techniques to the data set obtained by applying translation tools to existing processes, we intend to start a process that will end in an organology of digital instruments. Not only the processes themselves will be analyzed, but also the annotations made by users, as well as the automatic documentation previously obtained, and particularly the control parameters, names of input and outputs data, and dependencies. The relationship to works, where processes are in use, and metadata (authorship...) could also be analysed.

Conclusion

For preservation of digital material that is produced today, our approach is to use the tools that are available today, and particularly digital tools. To build a database is not sufficient, there is also the need of building knowledge out of it. To this end, we will use the most recent techniques in data mining and data analysis: neural networks, bayesian networks, or fuzzy logic. We will also validate the results obtained with these techniques by using statistical methodologies.
For building executable programs, we have to use the most recent techniques, for instance parallelism in order to get the best of multi-core processors. We have then to automatize the whole process, in order to save time and effort, including automatic generation of documentation, or automatic translation of objects from one expression to another one.
This does not exclude other approaches, and particularly those based on human reasoning and human activities. Our opinion is that our approach will give them some ways to explore, as well as a strong basis for experimentation and validation of new ideas.
Acknowledgements
This work is possible due to a grant of French National Research Agency, under the number 08-CORD-003.
We would thanks the ASTREE partners, and noticeally Yann Orlarey, Laurent Pottier and Pierre Jouvelot.

References

[1] Bernardini, N., Vidolin, A. 2005. Sustainable Live Electro-acoustic Music. In Proceedings of the International Sound and Music Computing Conference, Salerno, Italy, 2005.
[2] Assayag, G., « Problèmes de notation dans la composition assistée par ordinateur, », Rencontres Musicales Pluridisciplinaires. Musique et Notation GRAME, Lyon, 1997
[3] Merlier, B. "Vocabulaire de l'espace et de la spatialisation des musiques électro acoustiques : présentation, problématique et taxinomire de l'espace", Electroacoustic Music Studies Network – Beijing 2006
[4] Elleberger, E., "Notation symbolique musicale dans le domaine de la spatialisation", http://www.cmusge.ch/HEM/archivage/recherche/Spatialisation.htm
[5] Orlarey, Y., Fober, D., Letz, S., " An Algebraic approach to Block Diagram Constructions" Actes des Journées d'Informatique Musicale JIM2002, GMEM Marseille, p.151-158

Abstract
Libraries, archives, and museums (LAMs) have been the curatorial stewards of cultural heritage for some 5,000 years. The emergence of virtual worlds/immersive online realities as containers/conveyors of cultural heritage is presenting new preservation challenges to LAMs. Second Life is one virtual world that has achieved a level of ubiquity as to serve as model for all such digital environments, having a sufficiently large enough number of common problems from which to learn. The sheer complexity of the Second Life environment argues for a more "ecological" approach. The use of Second Life as an environment for government, commerce, education, and art has resulted in the creation of digital cultural heritage materials that are at risk. Successful approaches are examined, alternatives considered, and new directions are recommended.
Keywords: digital preservation, digital ecologies, digital cultural heritage, Second Life, machinima

Introduction

The virtual world Second Life is an example of why the library, archives, and museum (LAM) community needs to become more directly engaged in the decision-making processes of the digital preservation concerns surrounding virtual immersive environments. Taking the conservation of whole ecologies as a preservation strategy is appropriate to providing the context of meaning and relationship in Second Life that might otherwise be lost if current "data set" practices are maintained.

This departs from the usual concepts of digital preservation that focus upon the preservation of data as found in aggregate file structures or systems and their content. The latter, while capable of considerable automation, lends itself to the fragmentation of context and a susceptibility of losing the various relationships that may have initially existed between the aggregates of the data. Such losses are unacceptable and can be avoided by adopting more holistic approaches.

Definitions

Second Life is but one of many virtual worlds, yet it possesses qualities that make it a model for working with other similar immersive and interactive digital environments. It has the defining quality (for "virtual worlds") of what has been described as "persistent user-modifiable content"; this quality is best characterized as content which is created by users that persists within the virtual environment regardless of whether the creator/user is present and can be further modified by other users (Duranske, 2007) [1].

This is important, as it is an operative distinction from those games that utilize virtual, immersive environments. A virtual world may contain games (there are many games that are played in the Second Life environment), but games rarely achieve this status of "virtual world."

The Legacy

Libraries, archives, and museums (LAMs) have been curatorial stewards of cultural heritage for some 5,000 years. The nature of that stewardship has been evolving along with the available technologies. As our cultural manifestations have become increasingly digital in nature, LAMs have incorporated digital technologies to document, acquire, present, and preserve that culture. Documentation has become a part of the historian's domain and virtual worlds, while certainly au courant, have existed and flourished for a modestly significant amount of time while gaining the attention of historians who seek to record that emergence.

The definition of "significant time" is related to the rapidity with which digital technologies are embraced and then discarded. Complex environments were being created and used as early as 1998, in such projects as the Virtual UC Santa Cruz, which certainly foreshadows much of what we see today in Second Life. Perhaps not so coincidentally, Philip Rosedale founded Linden Lab in 1999, and for the first two years, the lab's work on Second Life was developing

it as an immersive, objective-driven game. From 2001 onward, the focus of Second Life development by Linden Lab has been user-generated content and a community-driven experience [2].

The Present

The emergence of virtual worlds as conveyors of cultural heritage is presenting new preservation challenges to LAMs. It is important to understand some of the distinguishing characteristics of what we mean by "virtual worlds." Virtual reality is no longer a new concept and the number of virtual worlds is growing regularly. There are whole "universes" of virtual worlds of varying complexity, depth, scope and breadth.

The highly social dimension that now exists in Second Life is of extreme importance. This increased "sociality" demands a critique of Second Life historiography; indeed, it requires more of an archivist's, an historian's, or an ethnographer's approach and sensibility to talk about how and what to preserve in Second Life, not the perspective of a computer scientist or game programmer.

It is important to distinguish between the different approaches needed to preserve game content versus virtual world content. Games, whether of the massively-multiplayer online type (MMOGs) such as Eve, World of Warcraft, or Everquest , the single player, or the various player versus player type, require different means of capturing the essence of the game, its story arcs, structures, and rules than when trying to capture the ecological complexity of a socially-driven immersive world. It is, however, difficult to decide the appropriate method for capturing MMOGs that approach the qualities that define a virtual world, as defined above. In such cases, a more ethnographic or documentalist approach is certainly appropriate in order to address the highly social nature of such games. Additional difficulties lie in Second Life's proprietary underlying environment, the user-created content with its issues of "ownership" and the currently "closed" nature (i.e., the rules, regulations, and requirements for entry and use of the software) of the environment.

The use of Second Life as an environment for government, commerce, and education raises legal and fiduciary issues. One has to consider if current records management practices are sufficient. A recent conversation with the founder of a group called "Archivists of Second Life" (an organization in Second Life having as a mission, among others, providing "leadership in the identification of records/archives of historical value to the residents of Second Life") revealed no encounters with anyone with a records management background [3]. Given the legal requirements and mandates usually associated with government, business, and education, the failure of Second Life to be on the proverbial radar of records managers is troubling at best and represents an area for further research, outreach, and education in itself.

The Future(s)

Linden Lab continues to evolve Second Life in new, but not totally unexpected, directions. The sheer scale of user numbers has drawn attention from the usual suspects, intent on making their fortunes. There is constant talk of the Lab selling or going public. Linden Lab has very deliberately announced various initiatives and strategic partnerships that suggest that they have an interest in promoting Second Life as a platform and technology for others to essentially license from the Lab. Distinction is now made between "Second Life Online Virtual World" and the "Second Life GRID Virtual World Platform" on the Second Life website [4]. The "business" of Linden Lab, or as it is increasingly self-identifying, "Linden Research, Inc.," is very deliberately appealing to enterprises, government, and education entities to use their platform. While this latter may be a shrewd business strategy, it carries considerable ramifications for any kind of preservation of content. Further, as these entities begin using the Second Life platform to conduct their business, there continue to be concerns about the records of transactions that occur.

That Linden Lab is diversifying their offerings should hardly be surprising in light of the competition from other companies and initiatives to provide similar platforms for virtual worlds. The Lab has released the Second Life viewing client (the "Viewer") to the open source community, but still maintains very tight control over the Second Life server code, as this is clearly seen by them as where money can be made. In a research partnership with IBM, experiments were conducted on 30 June, 2008, to see if alternative grids would be compatible to allow an avatar from the "official" Second Life grid to traverse to other grids not running on Linden Lab servers. The experiments were a qualified success in that the avatars were able to travel, but the "inventory" did not transfer across with the avatar.

The Open Grid Public Beta, as the experiment was known, was the attempt to develop virtual worlds with compatible underpinnings allowing inter-operability and new levels of customization, user control, or security. Along with the Open

Grid, there is a competing free open source initiative called the OpenSimulator that is quite forthright in attempting to create a virtual environment "similar to Second Life" [5]
More recently, Blue Mars, which is still in beta, is gaining attention, but fails to offer some of the inclusivity that is one of Second Life's hallmarks. By comparison, Second Life runs on Windows, Linux, and Mac OS X; Blue Mars is exclusively Windows. Blue Mars is gamer-oriented (reminiscent of the early incarnations of Second Life) and unlike Second Life, is not user-content driven. In direct competition with Second Life, Blue Mars is making an appeal to educators and business.

CONCLUSIONS

Digital preservation, at the most fundamental granularity, is rightly focused upon the preservation of "data" and there are best practices and guidelines in place for doing just that. But the preservation of digital culture, and especially virtual worlds, is more than simply "saving" data.
The preservation of digital culture must include and retain the context of that data which comprises the culture. This is an elusive goal and is at the root of the need for understanding that "the back up is not the archive." It is important to look at some of the "outside strategies" for more inclusive documentation and context-creation (Moser, 2009) [6]. This includes the utilization of open source solutions for institutions desiring the use of virtual worlds without tying them to the less-open environment of Second Life proper. Running one's own servers (especially Second Life viewer-compatible ones such as the aforementioned open source OpenSimulator), means being able to more confidently fulfill the obligations and legal requirements that may be incumbent.
A change in documentation is also needed. A more "ethnographic" approach, akin to archeology or cultural anthropology, is beneficial and appropriate for the documentation of digital culture. We are talking about whole systems, or preservation of an ecology, not "data set" preservation. The tools of the ethnographer and cultural anthropologist need to be adapted to use in Second Life. Documentary film in those fields has been highly effective; the application of this approach to Second Life is just beginning. Live motion screen capture and the use of machinima (i.e., in-situ created animated videos of the worlds) are providing documentaries that reflect the richness of the environment. This also includes the use of inworld tools that "follow" the avatar to record the important and elusive avatar-avatar interactions. No one tool is sufficient to capture this environment. We must consider the ecological approach, where each element has a direct connection to the whole. Such environments are multi-modal; our tools must likewise be multi-modal.

Bruce Damer is a pioneer in this approach. Damer has been documenting the history of virtual worlds, focusing upon environments with social interactivity as the emphasis, as opposed to the "gamer"-oriented ones. He has produced videos documenting some of the earliest worlds. His videos approach the subject from within their environments, allowing us a window into those worlds. As these predate the use of machinima, they do have some shortcomings not seen in Second Life machinima. They are still useful if only for the avowedly historical perspective utilized. His footage of Bonnie Devarco's Virtual UC Santa Cruz is an example succinctly documenting early virtual worlds.6

Damer's approach and work contrast sharply with the approach of his affiliate, Henry Lowood. Lowood is a key member of the "How They Got Game" collaboration at Stanford Humanities Lab, itself a part of the Preserving Virtual Worlds project, funded by the National Digital Information Infrastructure Preservation Program (NDIIPP) funded by the U.S. Library of Congress. Sadly, Lowood's approach conflates games with virtual worlds with less than satisfactory results. Virtual worlds simply are not the same as games. An example of this "game" approach to a virtual world, "Tabula Rasa: The Final Stand" is lacking in depth and context. The approach simply fails to capture the enormity of what is being examined. Tabula Rasa was a MMOG that was, by definition, a virtual world. What Lowood's approach has given us is snapshots when we need panoramic videos.

REFERENCES

[1] From the Editor: Are MMO Games "Virtual Worlds?" Benjamin Duranske, 25 February, 2007. http://virtuallyblind.com/2007/02/25/from-the-editor-are-games-virtual-worlds/ (Accessed, 15 October 2009)
[2] From the wiki, http://en.wikipedia.org/wiki/Second_Life … the entry has an overview of "what" Second Life is. It provides an excellent explanation of the breadth of experience to be found there. (Accessed 15 October, 2009)

[3] Taken from the “Archivists in Second Life” Group Charter, the mission statement says they exist to:
[4] *To promote the profession of records/archives preservation and records/archives access in and through Second Life.
[5] * To provide education, research and networking opportunities for archivists in and through Second Life.
[6] * To provide leadership in the identification of records/archives of historical value to the residents of Second Life.”
[7] This from the Linden Lab URL, http://lindenlab.com/, with its links to the online virtual world page, http://secondlife.com/, and the GRID page at http://secondlifegrid.net/. The GRID is described as a “Virtual World Platform for Business, Education, & Government | Second Life Grid”.
[8] The OpenSimulator project, while declaring itself to be alpha software, is compatible with the official Second Life viewer, runs on a variety of operating systems including various flavors of Linux, Mac OS X, Free BSD UNIX, and several versions of Windows. It is being promoted as a fully open source alternative to the closed source Linden Lab’s Second Life server.
[9] “The Avatar in The Archives: Issues of Documentation and Preservation of New Media Art and Virtual Worlds”, Dennis Moser, May, 2009. LIDA 2009 Conference Proceedings, LIDA 2009, Dubrovnik/Zadar, Croatia.
[10] The Internet Archives contains some 96 entries by Damer; while this is a keyword-driven count, most of those are linking to the videos from his collection. The Devarco video can be found, directly, at this URL: http://www.archive.org/details/vw_virtual-ucsc-devarco

Five years ago the Fondazione Rinascimento Digitale (http://www.rinascimento-digitale.it/) came into existence with the explicit task to promote the application of information and communication technologies in the field of cultural heritage. The young Foundation decided almost immediately to hold a conference. In this very same wonderful Teatro della Pergola (http://www.teatrodellapergola.com/) the Foundation organised a conference on 14-16 Dec. 2006 on the topics of access and preservation (http://www.rinascimento-digitale.it/conference2006.phtml). The conference looked at how new technologies were transforming knowledge and imposing new organisational requirements on our cultural institutions. Most of the papers were loosely classified as either on "digital libraries" or "digital preservation".

After the success of its first conference the Foundation has continued to work on projects in the broad fields of digital preservation, digital repositories, digital libraries and archives, and persistent identifiers.

Some 18 months ago it was decided to organise this second conference. Over the last 2-3 years much has changed. We saw IFLA (http://www.ifla.org/) in their 2009 conference in Milan focus on the future evolution of the library - and the way "libraries would drive access to knowledge". We saw ICA (http://www.ica.org/) becoming increasingly preoccupied with the challenges in exploiting new technologies to preserve "born digital" material. And we saw in the 2009 conference "Museums on the Web" (http://www.archimuse.com/mw2009/) major presentations on the institutional changes brought on by social media, on the creation of wiki communities, on digital asset management and digital preservation, on museum Web 2.0 sites, and on young audiences and creators.

So it was against this background that our conference title "CULTURAL HERITAGE: an active role for user communities" was conceived. We felt that the twin topics of access and preservation were just a valid as 3 years ago. However today it seems that users are not only able to adapt to technological changes faster than cultural institutions, but they are also driving innovation, becoming content producers and pushing institutions towards a new user-institution relationship. The Foundation was very fortunate to find support from the Italian Ministero per i Beni e la Attività Culturali and the US Library of Congress, and this produced a great cooperative effort in creating the sessions format, in providing speakers, in attracting high-quality papers and posters, and in promoting the event.

In addition to the support of many prestigious authorities and institutions, around 400 people attended the conference (including the pre- and post-conference tutorials). On the first day we were welcomed by representative from the Comune and Provincia of Firenze, the Regione Toscana, the Ente Cassa di Risparmio di Firenze (the parent organisation of the Foundation), the US Library of Congress, the Italian Ministero per i Beni e la Attività Culturali, and the European Commission.

We had 12 substantial invited talks on the state-of-art and state-of-practice in access and preservation. We had 24 papers presented in 2 parallel sessions: Digital library applications & interactive Web, and Sustainable policies for digital cultural preservation. And we had a poster session with another 11 papers. So a total of 47 speakers and presenters came together from 14 different countries. We saw speakers from major institutions such as the US Library of Congress, Instituto Centrale per il Catalogo Unico delle Biblioteche Italiane, the Italian Archivio di Stato, The British Library, the French Institut National de l'Audivisuel, the Austrian National Library, the Estonian National Archive, the European Digital Library Foundation and the European Commission. We also saw speakers from a multitude of prestigious academic institutions (North Carolina, Bath, Pisa, British Columbia, Helsinki, and Barcelona come to mind), and from major research labs. such as CNR, IRCAM, IBM and Max Planck.

So all the building blocks for success were present - a good cross section of high-quality cultural institutions and academic-research organisations presenting their activities and latest results.

Let us look more closely at the actual content. I want to retain our 2 traditional topics: access and preservation. My comments are largely based on the presentations and posters from the 2nd day. Many of the invited talks are in themselves masterful overviews of the state-of-art and/or state-of-play in specific fields of relevance to cultural institutions - and I

Chiara Cirinnà and Maurizio Lunghi (edited by), *Cultural Heritage on line: Empowering users: an active role for user communities. Florence, 15th - 16th December 2009* ISBN 978-88-6453-184-7 (print) ISBN 978-88-6453-187-8 (online)

would be doing the authors an injustice to try to summarise their contributions in a few lines. As such my comments should be read along with the collection of invited papers (and a longer version of these comments is available on the conference Web pages).
Equally I will not try to summarise all the papers presented. I have decided to pick out some points that I felt most relevant at the time. These are my personal comments and conclusions, and in no way do I intend to reflect negatively on the quality of the papers not mentioned. And naturally I hope I have understood and captured the key messages of the authors - and I here present my apologies if I mis-quote or mis-represent someones work.

FIRSTLY ACCESS

In this conference we saw two distinct trends concerning access. The first trend was towards very practical, large-scale Digital Libraries with an abundance of high-quality content being digitised and put online, and the second trend concerned a multitude of experiments with Web 2.0 technologies and social networks. Sitting between these two trends were a series of papers looking at the risks and benefits in adopting Web 2.0 technologies and social networking - and there were some concrete suggestions as to how to exploit the opportunities and manage to risks.

LARGE-SCALE DIGITAL LIBRARIES

How could I not but start with Jill Cousins of the European Digital Library Foundation who presented Europeana (http://www.europeana.eu) and Max Kaiser of the Austrian National Library who presented EuropeanaConnect (http://www.europeanaconnect.eu). Already today the prototype portal links to more than 5 million objects from more than a 1,000 European institutions and collections. And they promise 10 million items by 2010 and 25 million items by 2012. Europeana now has to integrate differing vocabularies, resources discovery tools, harvesters, metadata registries, and a multitude of licence agreements - then they want to add semantic processing, GIS- and time-related query options, etc. - and provide mobile access and on-demand ebooks. I admired the courage and optimism of Europeana - two essential qualities when trying to build a sustainable, large-scale pan-European Digital Library infrastructure and services.

Silvia Gstrein & Günter Mühlberger from the library of the University of Innsbruck described a trans-European ebooks on-demand network bringing together 20 libraries from 10 countries, including 6 national libraries (http://books2ebooks.eu). The approach taken gets users to co-fund the initial digitisation of rare or out-of-copyright material. Once the initial users demand has been met, the digitised book is made freely available to the public. What interested me was that a user survey indicated that around 70% of users were prepared to pay 50€ to get a copy of an out-of-print book.

And this is not all - over the 2 days we learned about a real abundance of high-quality content being put on line:

We heard Laura Campbell of the US Library of Congress mention the American Memory Website http://memory.loc.gov/ammem/index.html, which makes freely available more than 14 million historical primary source materials. In addition we also have the 200,000 documents in the Global Gateway (http://international.loc.gov/intldl/intldlhome.html) and the 100,000 newspaper pages in the US National Digital Newspaper Program (http://chroniclingamerica.loc.gov).
On the 1st day Daniel Teruggi mentioned that French Institut National d'Audiovisual provides access to more than 100,000 documents and more than 5,000 hours of audio-visual material.
Thomas Kirchhoff and his co-authors from the museum information systems group in Konstanz University presented BAM the German cultural heritage portal for libraries, archives and museums (http://www.bam-portal.de). This is another major portal effort to provide online access to 41 million records in the form of catalogues, repertories and inventories.

Lauri Leht from the National Archives of Estonia (http://www.ra.ee) talked about digitising around 5 million images, most from church books, and also putting online all 8 million archival heading thus allowing users to avoid searching through paper records.

Aly Conteh from the British Library & Asaf Tzadok from IBM in Haifa used the example of the National Library of Australia's newspaper Website (http://newspapers.nla.gov.au/ndp/del/home), which today has put 104,000 articles online.

Andrea Fojtu of the Czech National Digital Library (http://www.ndk.cz/project/view?set_language=en), discussed their plans to digitise 1.2 million documents or 350 million pages over next 20 years.

Christoph Müller from the Ibero-American Institute in Berlin (http://www.iai.spk-berlin.de/) talked about their plans to put online around 1 million items (plus 900,000 press clippings, 200,000 microforms, more than 70,000 maps, etc.).

Friederike Kleinfercher & Kristina Koller from Max Planck Digital Library looked at a development within the eSciDoc portal (https://www.escidoc.org). They mentioned that their ViRR project (http://test-virr.mpdl.mpg.de:8080/virr/) contains about 20,000 scans of legal artefacts from the Holy Roman Empire.

PORTALS, QUALITY CONTENT, USAGE AND SUSTAINABILITY

In looking at the emergence of these large-scale digital libraries we can see a move by cultural institutions towards services based upon portals. They want to present their content in a more user-friendly way, to offer new levels of interactivity, and to introduce user-oriented services sometimes working together with specific communities of interest. The list of digital library initiatives given above is just the tip of the iceberg - there are hundreds of other large- and small-scale projects being planned and implemented around the world. Yet in talking with authors most felt that that authentic high-quality content was still lacking on the Web. This means that institutions still saw a need to continue to digitise and expand their online digital collections.
Yet I was worried by the absence of real data on the actual usage of the services already available. And I also missed a real discussion on the sustainability of the investments already made.

BUT WHAT ABOUT STANDARDS?

Standards were not a specific topic in many of the papers presented at this conference, nevertheless the impression was that we have moved from a situation of uncertainty (a seen in the 1st conference in 2006) to one of relative clarity and even apparent abundance - few authors mentioned the lack of, or complexity of, todays standards as a barrier or risk.
In my opinion this may not be the true situation. Firstly this conference was associated with pre- and post-event workshops on long-term preservation and Dublin Core. Both offering ample opportunities to focus on standards such RDF and metadata, digital formats, as well as tools, practices, and approaches to risk management, etc. More importantly some authors hinted at the fact that smaller institutions appear still to have a very naive approach to standards for both digitisation and long-term preservation. I still think that there is a place for standards development and promotion through very practical guidelines, trails and experiments. Equally the success of the pre- and post-conference workshops shows, if anything, an increasing need for tutorials and training courses.
On the other hand we heard from several authors that one of the main advantages of Web 2.0 as a technology platform is that it is an existing (at least semi-standardised) infrastructure and is cost-effective for institutions, even if the different social networks are not interoperable (Kelly & Oppenheim). I might add that the novel integration of RFID tags, GPS and semantic Web by Kauppinen (representing the consortium behind the SMARTMUSEUM project http://smartmuseum.eu) also has the advantage of adopting what are becoming cheap and well-defined industrial standard components.

WEB 2.0 TECHNOLOGIES: RISKS & BENEFITS

Looking beyond these large digital library efforts we can see more experimental projects exploiting Web 2.0 technologies and social networking as a way to involve users in the creation and maintenance of distributed collections of cultural material.
Smiliana Antonijevic of the Royal Netherlands Academy of Arts and Sciences working with Laura Gurak from the University of Minnesota looked at trust in online interaction. They noted that modern-days users want fast, accurate systems that have some embedded intelligence and can be customised. However they also noted (and I think rightly so) that above all users want systems that are trustworthy. Some of todays digital repositories are certainly moving in that direction - becoming reliable and persistent over time and engendering trust with their users. On the positive side, involving users can transform a static and historical content authority into a dynamic, multi-faceted and evolving body

of localised knowledge. The down side is that information provided by users can be incorrect, incomplete, misleading, corrupt or highly biased. The authors called for cultural institutions to understand how to protect their status as trusted authorities and to learn how to transfer trust to external sources of information. However the authors also noted that the main socio-cultural features of trust have remained stable over the years and much can be learned by harvesting results published over the past 30 years.

Brian Kelly from UKOLN & Charles Oppenheim from Loughborough University echoed the point of view expressed by Antonijevic & Gurak, but they went one step further. On the one hand they recognised that Web 2.0 concepts were moving into the cultural institutions (they mentioned Library 2.0 as being an accepted expression - it even has its own wiki entry at http://en.wikipedia.org/wiki/Library_2.0, and Archive 2.0 and Museum 2.0 being not far behind). On the other hand they also extended the list of concerns and risks: services may not be secure, reliable or interoperable, they are open to misuse, and there are still open legal issues concerning the relationship between cultural institutions and users as content providers (and there are also outstanding copyright issues, the risk of misleading or inaccurate information, a failure to respect data protection laws and personal privacy, and the ever-present risk of posting illegal content). On the positive-side the authors noted that social networks are popular and easy to use, they can engage with new user communities, and are cost effective because they exploit an existing infrastructure. So the real issue is to help cultural institutions learn how to manage the risks involved in building and exploiting social Webs. The authors went on to propose a risk/benefits framework where institutions should be explicit about intended use, benefits, risks, miss opportunities when not adopting a new technology as well as the costs when adopting it, how to minimise risk, and the need to clearly document the evidence used in the analysis.

There were three further papers that highlighted the difficulties in understanding and meeting user needs. The first paper was by Alida Isolani from Scuola Normale Superiore di Pisa looking at empowering users without weakening the digital resources. The authors provide a collection of Renaissance texts for humanities scholars (http://bivio.signum.sns.it). The contents are valuable and regularly accessed by academics, but the advanced retrieval tools available are not well used - and users tend to access the site in a "traditional way". The authors concluded that the tools need to be simplified - by making them more complex! More services have to be offered (analysis, note, mark, correct, edit, etc.) and more formats have to be supported. It will be interesting to see if this approach really increases user demand. The second paper was from Fred Stielow of the American Public University system looked at community building in an online university context. The author extended the list of very practical risks/problems he faces daily - ranging from the problem of price negotiations in todays chaotic rights marketplace, through the need to keep costs down when creating metadata and catalogues, to ways to improved tailoring for individual students. The third paper in this group was from Jeremy Hunsinger of Virginia Tech., who looked at the problem of usage of an event-driven memory-bank (http://www.april16archive.org). The author reminded us that the memory bank is a collection (or memorial) of digital artefacts contributed after the April 16 tragedy at Virginia Tech. (where 32 people were killed). He mentioned that today his real problem is now a lack of visitors or users, and the author asks "what happened to the users?" without really being able to find an answer!

Looking beyond the risks: practical experiences

So a risk-benefit analysis is an absolute must, but there are still many ways to exploit Web 2.0 and social networking, keep the risks low, and obtain some valuable and practical results. Lets look rapidly at some practical experiences.

Cèsar Carreras & Frederica Mancini of the Universitat Oberta de Catalunya looked at Web production by small sized institutions. The authors discussed the aims and fears of institutions when faced with the Web 2.0 and the development of social networks. Users can express preferences and opinions, and this virtual community can represent a new life for a small institution (encouraging physical visits, promoting daily discussions, creating empathy with "friends of the museum", stimulating user content production and commentaries). But how to do this properly? There are risks: sterile tools that create nothing new, alienation of the users through abusive advertising, deforming the institutional identity, etc. In concluding the authors discussed the different approaches. The key for a small institution appears to be to create

a local community of interest that supports not only content creation but also has reliable elements of content quality checking and validation (through local professionals, teachers, etc.). The authors stressed that quality checking is an expensive process for a small museum, and yet poor quality content can undo all the benefits that a institution creates in its local community.

Lauri Leht of the Estonian National Archives (http://www.ra.ee) looked at involving users in enriching digitised archival material. They have digitised around 5 million images, and put online all 8 million archival heading. Archive volunteers have been employed doing quality checking, helping to understand the content and describe the content in a structured way, and in collecting similar data from different archival sources (remembering that documents are in Estonian, German and Russian with differing alphabets and full of errors).

Aly Conteh from the British Library & Asaf Tzadok from IBM in Haifa talked about ways to foster user collaboration during mass digitisation - using as an example the National Library of Australia's newspaper Website (http://newspapers.nla.gov.au/ndp/del/home) which supports collaborative correction of Optical Character Recognition (OCR) output. Generally speaking over 20% of the text of an early 19th century newspaper will not be correctly recognised (and this is equally true of many types of historical text). In-house checking and re-keying is not a valid option when digitising millions of pages. The authors argued for collaborative user correction and validation to improve the accuracy of OCR results. And improved OCR means better text mining, resource discovery, and overall accessibility. Already this newspaper project, in its first 6 month, brought together nearly 3,000 people to correct 104,000 articles.
The authors concluded by suggesting a hybrid approach: improved OCR technologies for automated text recognition linked with collaborative correction (not just correcting errors but also helping to train and enhance the OCR's engine vocabulary and language analysis features).

TOP PROBLEMS: COST, EXPERTISE, AND INFORMATION MANAGEMENT

Before moving on the topic of preservation, I would like to close this section on access by referring to the paper of Wendy Duff and co-authors from Toronto University. They looked at the impact of new technology on the museum environment in the US - and based their discussion on semi-structured interviews with 16 US-based senior museum professionals. The 3 most common challenges facing those interviewed were: cost of designing, implementing and maintaining technology, the lack of in-house expertise, and information management.
The starting point is a series of bold quotes saying that "a museum without a collections database and a Web presence is hardly considered professional" and that museums are moving from "object-centred to experience-centred design". However on the down side "not all institutions are using online access equally well" and many funding agencies and museums professional don't fully understand the challenges IT poses for museums.
Finding from the interviews included:
There was no consensus about the extent to which the core-mission of museums has been impacted by new technologies (have they changed the core mission of the museum, does it help to attract a broader audience, or connect better with the local community, or change the way the museum works, or has it altered the way a museum sees itself, ...).
However most agreed that museums have been physically transformed by the proliferation of new technologies (multimedia installations have changed the way exhibitions are held, changes made in dissemination and collections management, for some interviewees 3D imaging is become an increasingly important tool, technology also helps professionals remain curious and creative in developing their expertise and plans for the future, and technology is now an essential tool in linking objects with the information about them, even if the management of legacy data remains a challenge).
For the majority of interviewees the major challenges are: cost of designing, implementing and maintaining technology, the lack of in-house expertise, and information management. Databases need to be created, data needs to be migrated and cleaned, metadata created and maintained, vocabularies need to agreed upon and shared, ..., and all this takes time, expertise and is expensive. Despite some people being technologically savvy, the majority of people were seen as not computer literate, so high-tech services might be simply an over-kill. Some people noted that poor quality or out-of-date information distributed over the Web can reflect negatively the reputation of the museum and its staff. And many museums don't understand fully the cost/benefit of introducing new technologies.

An important point made by the authors was that many museums don't appear to be dealing in the most efficient and cost effective manner with long-term digital preservation, e.g. digital photos are just being dumped on CD-ROM's and stored on shelves.

And now Preservation:
In this conference our aim was also to review progress in digital preservation technologies and applications (and we should not forget that there was a pre-conference event dedicated to the basic concepts and practices of long-term digital preservation).

Sven Schlarb from the Austrian National Library (and his co-authors from The British Library and ARC - the Austrian Research Centers) looked at the Planets Testbed (http://www.planets-project.eu) which is a Web-based application that provides a controlled collaborative environment for scientific experimentation in digital preservation. The authors outlined how the testbed was used, how a tool was tested and assessed, and how the results analysed. Tools can be compared, preserved objects can be validated, emulation experiments performed, and a preservation plan created with recommendations. There is already a community of users sharing the experiments and a lot has been done to provide access to results (preservation services are offered, annotated datasets are available, validation services can check for valid and invalid document types, etc.). The authors closed by noting that the testbed will soon be a freely available public service.

Sam Coppens and co-authors from Ghent University looked at digital preservation using a semantic metadata schema of the PREMIS 2.0 preservation standard (http://www.loc.gov/standards/premis/). The authors kicked-off with an impressive list of all the different types of metadata that are needed: descriptive for search and general archive management, binary to describe the bitstreams, technical describing the files, structural for the representation information, preservation indicating provenance, context, etc., and finally rights metadata. The authors have extended PREMIS 2.0 to include the legal roles that people, organisations or software application can have. In concluding the authors stated that employing a 2-layer model allows the upper level with descriptive metadata to be made public (rights permitting), whilst the lower level with the legal roles remains in the hands of the institution.

Christoph Müller from the Ibero-American Institute (and his co-authors including from IPK-Fraunhofer) looked at user demands and preservation requirements for digitisation. The Institute in Berlin (http://www.iai.spk-berlin.de) is Europe's largest special collection on Latin America, Portugal, Spain and the Caribbean. The paper looks at the differing, often conflicting, requirements of scholars, librarians, and users. For example scholars want digitised copies to be as authentic as possible and tend to focus on making rare and unpublished material available. Librarians want digitisation to integrate well into their workflow and enable automated quality controls and indexing during scanning. Users (students, public, etc.) want content and context, want full-text search, and want want fast and easy access (in particular for exam preparation). The authors now have a "wish list" for features of a future digitisation system, starting with flexible automated digitisation, then interactive quality control, excellent picture quality, easily generated metadata, etc.

Andrea Fojtu and co-authors looked at long term preservation in the Czech National Digital Library (http://www.ndk.cz/project/view?set_language=en). The authors discussed their digitisation and long-term preservation objectives (e.g. digitisation of 1.2 million documents or 350 million pages over next 20 years using robot scanners). They rightly identify the organisational challenges as being as important as the technical issues (nice expression "institutions must be ready for a business change, well before the scanners produce the first pages"). The authors provided a long list of practical suggestions, ranging from the creation of a digital preservation department to the changes needed in existing in-house workflows and the relocation and retraining of staff.

More on metadata!
Thomas Risse from the L3C research centre (and co-authors from the European Archive, the Hungarian Academy of Science, and the Max-Planck-Institut für Informatik) looked at how to turn stored Webpages into a living Web archive. The authors started by noting that Web archival has value (for scholarly studies, market analyses, IPR disputes, etc.),

and there are now emerging industrial services in addition to the usual library and archival organisations. However Web content is highly dynamic, volatile, and in many formats. In addition physical media decays, technologies become obsolescent, encoding standards change, authenticity and integrity are difficult to maintain, etc. To go beyond just "freezing" Web pages, the authors looked at archival fidelity (capturing also the hidden and social Web, but not spam), coherence (identifying, analysing and repairing temporal gaps), and interpretability (ensuring accessibility and usability of the archive including the evolution of terminology, etc.). The authors discussed 2 applications: a "social Web archive" (for dynamic and varied user interactions) and a "streaming archive" (for audio-visual content) - all within a EU-funded project called LiWA (http://www.liwa-project.eu).

Felix Engel from FernUniversität Hagen (and his co-authors from Deutsche Nationalbibliothek and the company GLOBIT) looked at context-oriented scientific information retrieval with the specific aim to enable reuse of scientific publications, data and multimedia objects. This requires the capture and storage of additional metadata during all life-phases of the digital object, before, during and after archival. As noted by the authors this supports the goal of digital preservation by enabling reuse (and without being able to contact the object creators). Thus born-digital objects are defined not only as themselves, but also by life-cycle processes such a creation, appraisal, archival and adoption (unpacking, ingestion, adaption, transformation, display, emulation, access, aggregation, contextualisation, etc.) and reuse (including updates to the metadata).

Maristella Agosti and her co-authors from the University of Padua looked at cross-language access to archival metadata. The authors argued for an approach that would allow archival metadata to be both easily machine processable and permit cross-language solutions developed in the library community to be easily adopted by archivists.

Going beyond "conventual" digital preservation

Jerome Barthelemy from IRCAM in France (and his co-author including from McGill University) looked at real-time audio processing and a notation for contemporary music. The authors want not only to preserve music but also preserve the ability to re-perform the works live, e.g. for modern interactive works that are today completely dependent on a specific hardware and software implementation. They claimed that it is necessary but not sufficient to simply record and preserve outputs. The actual hardware and software used (called a patch) to process the input (from the performer or pre-recorded) must also be preserved. An alternative might be to develop a emulator, but this looks to be fraught with difficulties and uncertainties. Migration, moving from one technical environment to another, has its place. However the authors put forward the idea of virtualisation, or describing the electronic modules employed using abstractions. So a representation of signal processing modules can be found to describe say a violin played together with live electronics, and this can scored alongside the instrumental parts. Now comes the issue of musical notation.

Dennis Moser from the University of Wyoming looked at conserving digital ecologies such as Second Life. The author premiss was that our libraries, archives and museums will need to preserve complex environments such as Second Life (http://secondlife.com/ a user-generated and community-driven "experience"). He argued that it is inappropriate to simple store files, loosing the relationships that existed between aggregates of data. Massive-multiplayer online games can pose problems when trying to capture the stories, structures, rules, etc. and the complexity is increased by the closed proprietary environment used in Second Life and with the user-generated content that has separate ownership. Moser argued for a more "ethnographic" approach when dealing with worlds such as Second Life, i.e. preserving an ecology rather than a data set. He suggests that producing video documentaries, with for example machinima (http://www.machinima.com), can go some way to capturing what actually happens inside Second Life.

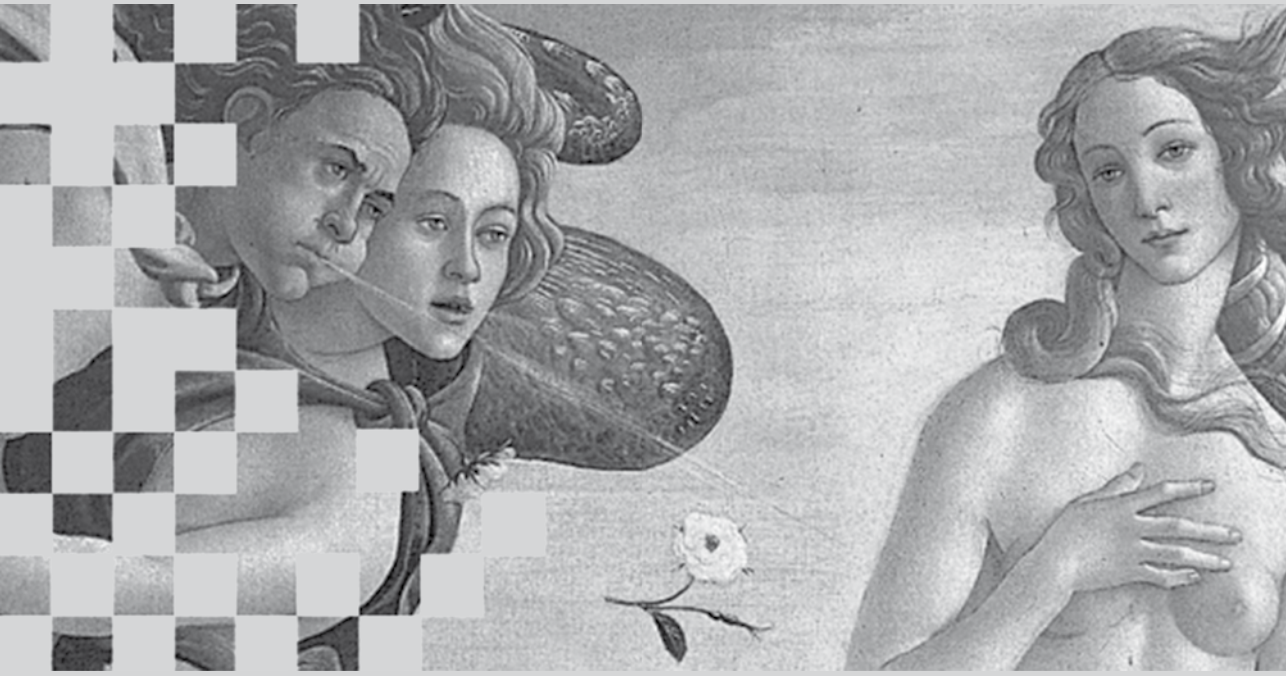
A need to combat fragmentation in long-term digital preservation work

More generally the different papers and presentations on digital preservation highlighted the complex nature of the problem. In particular when dealing with environments that change and evolve in a disordered and quite rapid way.
We saw in some papers tools being developed that look to be based upon self-defined principles and methods, but which are specific to individual sectors.

The risk is fragmentation. Different ideas and approaches can rapidly lead to isolation and dead-ends when set against a rapidly changing technological and organisational landscape - even more so when users look to be driving innovations.
What we need is to share results and experiences in a cross-domain confrontation. We need to promote a common understanding of the different scenarios and frameworks that underpin efficient digital preservation policies. We need a set of networks (regional, national, European) to:
Avoid useless duplication and foster a single-minded concentration on the long-term sustainability of approaches;
To test research results (to breaking) across a set of complex, cross-disciplinary tasks;
To offer high-quality training/educational events with a focus on real-world problems and using real content.

From "what might be" to "what is"

At the start of these conclusions I mentioned the objectives of the Fondazione Rinascimento Digitale, but today the real question is concerns what we can expect from the Foundation in the coming years. I personally think that the key will be to make its research, training courses, workshops, and above all its results as relevant as possible to cultural heritage professionals and academics. But to do so it will need feedback - positive and negative. Please go to the Foundations Website - look at the results, use them, adopt them, and tell the Foundation what you think. It needs constructive criticism in order to progress. And suggest to the Foundation what you think it should be doing next.
But criticism is not enough, it needs also to be congratulated when it has done something positive. And I think this conference is a positive result. What we have seen over the last 2 days has been less to do with "what might be" and more to do with "what is" - that is real-world considerations on building large digital collections, the practical reality in working with Web 2.0 technologies, the risks and benefits in working with users within large social networks, and the state-of-play in long-term digital preservation.
For making this conference happen our thanks must go to the Fondazione Rinascimento Digitale, and to the Italian Ministero per i Beni e la Attività Culturali and the US Library of Congress for the support they provided. In addition we had an impressive list of sponsors: the Ente Cassa di Risparmio di Firenze (the parent organisation of the Foundation), the Comune and Provincia of Firenze, the Regione Toscana, and UNESCO. and an equally impressive list of supporters: CNR, W3C, Liber, IFLA-PAC, European University Institute, europeana, Planets, CIVITA, and many more.
Our thanks must also go to the authors, speakers, and session chairs for providing the content of our conference and for making it such an intellectually stimulating event. Equally our thanks go the all the participants who attended all the sessions, asked questions, created debate, and who made this conference so dynamic and - in many ways a real, tangible, albeit "old-fashioned" social network.
As a final comment and with the desire to build on the embryonic community created over the 2-day conference I ask the organisers to:
Post on the conference Webpage a simple link-page listing all the links mentioned in the all the different papers, posters, and presentations (pointers to collections, tools, projects, etc.);
Send out a questionnaire to all attendees asking for comments concerning the conference content and organisation;
Consider ways to build on the community spirit established over the 2 days through a short regular newsletter or even a dedicated Facebook page (the approach must be validated with the user community).



Papers accepted

Abstract

In this paper we analyze the ratio between Digital Library (DL), archives and multilingualism. We focus our attention on the interoperability issues that need to be faced when you attempt to make different cultural institutions cooperate, to allow a selective and pinpoint online access to their resources, and to enable cross-language retrieval of their materials.

Introduction

Digital Library (DL) systems have been becoming the fundamental tool for managing, exchanging and searching cultural digital resources and as a research field has seen continuous growth over the last ten years. The central role of DL in fostering access to our cultural heritage is also enhanced by the European Commission which financially supports many projects related to DL, such as the TELplus project[1], which aims to offer a free service to access the resources of the 48 national libraries of Europe in 20 languages, or the Digital Repository Infrastructure Vision for European Research (DRIVER) project[2], the goal of which is to develop a pan-European Digital Repository Infrastructure by integrating existing individual repositories from European countries and developing a core number of services, including search, data collection, profiling and recommendation. Furthermore, the "European Commission Working Group on Digital Library Interoperability has the objective of providing recommendations for both a short term and a long term strategy towards the setting up of the European Digital Library as a common multilingual access point to Europe's distributed digital cultural heritage including all types of cultural heritage institutions" [4]. In particular, the recipient of these recommendations is Europeana[3] , which aims at addressing the interoperability issues among European museums, archives, audio-visual archives and libraries for the creation of the "European Digital Library". From this picture we can see that DL are not merely the digital counterpart of traditional libraries, but they are the fundamental tool for pursuing interoperability between different cultural organizations such as libraries, archives and museums. Collecting and managing the resources of these organizations is fundamental for providing wide, distributed and open access to our cultural heritage.

Currently, libraries are the foremost components of DL, this is due to the availability of technologies well-suited for them and that have been adopted by DL since their conception such as the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) that is the standard de-facto for metadata exchange in distributed environments and the Dublin Core[4] (DC) metadata format which is a tiny and lightweight metadata format that is getting the preponderant mean to exchange information. Archives and museums should adopt these technologies to exploit the services offered by the DL systems; two European projects pursue this goal: the APEnet[5] (Archives Portal of Europe on the Internet), which aims to build an Internet Gateway for Documents and Archives in Europe, and the Athena (Access to cultural heritage networks across Europe) project[6], which aims to reinforce, support and encourage the participation of museums and other institutions coming from those sectors of cultural heritage not fully involved yet in Europeana. Unfortunately, the process of adopting these technologies and exploiting the DL system advanced services is not as straightforward as it is for the libraries; this is due to the nature and the organization of the archives and of the museums as cultural institutions. In this paper we shall concentrate on archives because the problematic issues of museums can be related to those of archives; indeed, often museum resources are described and organized as archival resources. The archival structure

[1] http://www.theeuropeanlibrary.org/telplus/
[2] http://www.driver-repository.eu/
[3] http://www.europeana.eu/
[4] http://www.dublincore.org/
[5] http://www.apenet.eu/
[6] No Website yet available.

Chiara Cirinnà and Maurizio Lunghi (edited by), *Cultural Heritage on line: Empowering users: an active role for user communities. Florence, 15th - 16th December 2009* ISBN 978-88-6453-184-7 (print) ISBN 978-88-6453-187-8 (online)

is deeply hierarchical and the relationships between the documents must be retained to express their full informational power. These characteristics lead to the development of metadata standards such as the Encoded Archival Description (EAD) which are not particularly well-suited to be used within the DL systems. These standards may be a barrier towards the interoperability between the cultural institutions and towards the automatic processing of the data. These difficulties have moved archives away from full participation in DL, in particular they have limited the access to several services offered by DL systems. For both archives and libraries, multilingual access to the resources is a key point especially in the European context; indeed, multilingualism also promoted the CACAO European project[7] which aims to offer an innovative approach for accessing, understanding and navigating multilingual textual content in digital libraries. Furthermore, the CACAO infrastructure will be adopted by "The European Library" to promote aggregation of different contents at the European level. In this paper we analyze the problematic issues which could prevent the use of the multilingual services within the archival digital resources. Moreover, we shall propose a methodology that permits us to exploit the techniques adopted by the libraries with the archival metadata, enabling a multilingual access to these valuable resources.
The paper is organized as follows: Section 2 introduces the three main techniques to address metadata-related challenges in a multilinguistic environment. In Section 3 we briefly describe the archival organization and we explain why EAD metadata format does not work well in distributed and multilingual environments. In Section 4 we present our methodology which maps the EAD files into a combination of sets and DC metadata enabling the use of the cross-language techniques. Finally, in section 5 we draw some conclusions.

Cross-Language Access: Metadata-Related Challenges and Solutions

In the European Union (EU) there is a huge need to provide cross-language access to information; this is due to the diversity and multilingual EU environment where there are 23 official languages spoken in 27 member states. Cross-language access to information leads to problems of both semantic and syntactic interoperability [6]. Many solutions such as those adopted by the CACAO Project aim to address these problems mainly through the use of metadata, which provide access to a multilingual corpus of cultural resources.
A system which has to provide cross-language access to information must address two important metadata-related challenges which can be tackled by specifying the language of the metadata fields [6]: false friends and term ambiguity. To address these issues three main solutions are usually considered:

- Translation: A query formulated in the user language is automatically translated in the other supported languages and then submitted to the system. This solution is not free from the false friends issue.
- Enrichment of Metadata: The aim is to make the intended meaning of information resources explicit and machine-processable, to allow machines and humans to better identify and access the resources. The language would thus be provided in the metadata itself.
- Association to a Class: Terms are associated to a fairly broad class in a library classification system such as the Dewey Decimal Classification (DDC). This is a common solution for the term ambiguity problem and is similar to synsets used in WordNet[8].

The specification of the language of metadata field enables the full exploitation of metadata for cross-language purposes. If metadata do not come with or cannot be enriched with the language of the field, it is useful to rely on the association to a class technique. This useful technique relies on the use of the subject field of metadata; it is not always possible to determine the subject of a metadata or of a term. This is particularly true for archival metadata where determining the subject can be very difficult.

Archival Metadata and the EAD Format

An archive is a complex cultural organization which is not simply constituted by a series of objects that have been accumulated and filed with the passing of time. Archives have to keep the context in which their documents have been

[7] http://www.cacaoproject.eu/home/
[8] http://wordnet.princeton.edu/

created and the network of relationships among them in order to preserve their informative content and provide understandable and useful information over time. The context and the relationships between the documents are preserved thanks to the strongly hierarchical organization of the documents inside the archive. Indeed, an archive is divided by fonds and then by sub-fonds and then by series and then by sub-series and so on; at every level we can find documents belonging to a particular division of the archive or documents describing the nature of the considered level of the archive (e.g. a fond, a sub-fonds, etc.).

The union of all these documents, the relationships and the context information permits the full informational power of the archival documents to be maintained. In the digital environment an archive and its components are described by the use of metadata; these need to be able to express and maintain such structure and relationships. The standard format of metadata for representing the complex hierarchical structure of the archive is EAD [7], which reflects the archival structure and holds relations between documents in the archive. In addition, EAD encourages archivists to use collective and multilevel description, and because of its flexible structure and broad applicability, it has been embraced by many repositories [7]. The use of EAD is widespread in the United States of America and also in the EU; for instance the "Nationaal Archief"[9] in the Netherlands preserves a big collection of EAD metadata in Dutch or the "Archives Napoleon"[10] is based on EAD metadata in French. It is important to include archival metadata in DL because they retain unique and valuable information and at the same time it is very useful to enable multilingual services to access and retrieve them.

Unfortunately, the structure of EAD turns out to be a very large eXtensible Markup Language (XML) file with a deep hierarchical internal structure. On the other hand, EAD allows for several degrees of freedom in tagging practice, which may turn out to be problematic in the automatic processing of EAD files, since it is difficult to know in advance how an institution will use the hierarchical elements. The EAD permissive data model may undermine the very interoperability it is intended to foster. Indeed, it has been underlined that only EAD files meeting stringent best practice guidelines are shareable and searchable [10]. Moreover, there is also a second relevant problem related to the level of material that is being described. The EAD schema rarely requires a standardized description of the level of the materials being described and this possibility is often ignored, as pointed out by Pitti in [7]. Therefore, the access to individual items might be difficult without taking into consideration the whole hierarchy. This issue compromises the possibility of automatically enriching the metadata for multiliguality purposes. A single EAD metadata is used to describe an entire archive, thus in a single metadata we can find very different subjects. With this organization it is very difficult to disambiguate the terms or to identify the subject of metadata; with the EAD metadata the "association to a class" solution is essentially unworkable. Moreover, sharing and searching archival description might be made difficult by the typical size of EAD files which could be several megabytes with a very deep hierarchical structure. Indeed, each EAD file is a hierarchical description of a whole collection of items rather than the description of an individual item. On the other hand, users are often interested in the information described at the item level, which is typically buried very deeply in the hierarchy and might be difficult to reach.

A Methodology to Enable Both Cross-Language Access and Exchange of EAD Metadata

In [2] a solution was proposed to enable the sharing of EAD metadata in a distributed environment and enabling the variable granularity access to the data; this solution maintains also the integrity and the structure of the described archive exploiting OAI-PMH inner structure and the DC metadata; indeed, it is based on a methodology which enables an EAD file to be represented as a combination of OAI-sets and several DC metadata. To properly understand this methodology it is worthwhile briefly describing the functionality of OAI-PMH called selective harvesting and how its internal organization based on OAI-sets can be used to express a hierarchical structure as an organization of nested sets [3].

Selective harvesting is based on the concept of OAI-set, which enables logical data partitioning by defining groups of records. Selective harvesting is the procedure which enables the harvesting only of metadata owned by a specified OAI-set. In OAI-PMH a set is defined by three components: setSpec which is mandatory and a unique identifier for

[9] http://www.nationaalarchief.nl/

[10] http://www.archivesnationales.culture.gouv.fr/chan/chan/archives_napoleon-averti. htm

the set within the repository, setName which is a mandatory short human-readable string naming the set, and setDesc which may hold community-specific XML-encoded data about the set. OAI-set organization may be hierarchical, where hierarchy is expressed in the setSpec field by the use of a colon [:] separated list indicating the path from the root of the set hierarchy to the respective node. For example, if we define an OAI-set whose setSpec is "A", its subset "B" would have "A:B" as setSpec. When a repository defines a set organization it must include set membership information in the headers of the records returned to the harvester requests. We exploit this structure to represent a hierarchical structure such as a tree data structure as an organization of nested sets as shown in Figure 1. Here we can see that each node of the tree can be mapped into a set, where child nodes become proper subsets of the set created from the parent node. Every set is subset of at least one set; the set corresponding to the tree root is the only set without any supersets and every set in the hierarchy is subset of the root set. The external nodes are sets with no subsets. The tree structure is maintained thanks to the nested organization and the relationships between the sets are expressed by the set inclusion order [3]. This methodology allows us to decompose the EAD tree structure into an organization of OAI-sets where the elements belonging to a set are metadata records. The structure of the EAD is maintained by the OAI-sets and the data are mapped into many DC records. As far as the mapping of the actual content of EAD items into DC records is concerned, we adopt the mapping proposed by Prom and Habing [9]. Our solution differs from [9] from a syntactic point-of-view: we propose to maintain the hierarchical structure of EAD throughout an organization of OAI sets containing the DC records mapping the content of EAD items. In [9] the hierarchical structure is maintained by means of several pointers connecting the DC records to the original EAD file.

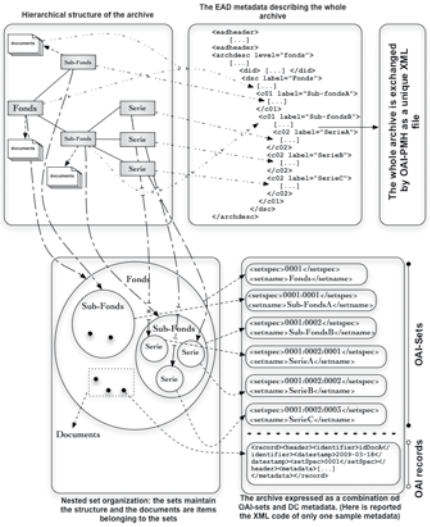


Figure 1 An EAD file mapped into a collection of OAI-sets and DC metadata records.

In Figure 1 we can see two approaches to representing the archival organization and documents. The first approach is the EAD-like one in which the whole archive is mapped inside a single XML file. All information about fonds, sub-fonds or series as well as the documents belonging to a specific archival division are mapped into several XML elements in the same XML file. With this approach we cannot exchange precise metadata through OAI-PMH, rather we have to exchange the whole archive. At the same time it is not possible to determine a specific subject or to access a specific piece of information without considering or accessing the whole hierarchy.
By means of our approach, which graphical representation is shown in the lower part of Figure 1 we can transform archival metadata into a collection of DC metadata and OAI-sets. This solution is particularly well suited for use in the context of the several European projects and in particular for the CACAO project which relies on OAI-PMH to

harvest the metadata and on DC records as minimum metadata requirement. In this way the solutions proposed to enable cross-language access to digital contents can be applied also with the archival metadata opening these valuable resources to a significant service offered by the DL technology. Indeed, the decomposition of an archive from a single EAD file into several DC metadata makes it easier to determine the subject of each single metadata and thus to apply the "association to a class" solution; in the same way the metadata enrichment can be adopted because the DC metadata are well-suited to automatic processing. As we can see, thanks to this methodology, the cross-language solutions developed for the library context can be easily adopted in the archival context without any additional efforts.

ACKNOWLEDGMENTS

The work reported in this paper has been partially supported by a grant from the Italian Veneto Region. The study is also partially supported by the TELplus Targeted Project for Digital Libraries, as part of the eContentplus Program of the European Commission (Contract ECP-2006-DILI- 510003).

REFERENCES

[1] A. Bosca and L. Dini. CACAO Project at the TEL@CLEF 2008 Task.

[2] N. Ferro and G. Silvello. A Methodology for Sharing Archival Descriptive Metadata in a Distributed Environment. In Proc. 12th European Conf. on Research and Advanced Technology for DL (ECDL 2008), pages 268-279. Lecture Notes in Computer Science (LNCS) 5173, Springer, 2008.

[3] N. Ferro and G. Silvello. The NESTOR Framework: How to Handle Hierarchical Data Structures, In Proc. 13th European Conf. on Research and Advanced Technology for DLs (ECDL 2009), pages 215-226. Lecture Notes in Computer Science 5714, Springer, 2009.

[4] S. Gradmann. Interoperability of Digital Libraries: Report on the work of the EC working group on DL interoperability. In Seminar on Disclosure and Preservation: Fostering European Culture in The Digital Landscape. National Library of Portugal, September 2007.

[5] K. Kiesling. Metadata, Metadata, Everywhere - But Where Is the Hook? OCLC Systems & Services, 17(2), pages 84-88, 2001.

[6] B. Levergood, S. Farrenkopf, and E. Frasnelli. The Specification of the Language of the Field and Interoperability: Cross-Language Access to Catalogues and Online Libraries (CACAO). In Proc. Of the Int'l Conf. on Dublin Core and Metadata Applications 2008, pages 191-196.

[7] D. V. Pitti. Encoded Archival Description. An Introduction and Overview. D-Lib Magazine, 5(11), 1999.

[8] C. J. Prom. Does EAD Play Well with Other Metadata Standards? Searching and Retrieving EAD Using the OAI Protocols. Journal of Archival Organization, 1(3), pages 51-72, 2002.

[9] C. J. Prom and T. G. Habing. Using the Open Archives Initiative Protocols with EAD. In Proc. 2nd ACM/IEEE Joint Conference on Digital Libraries, (JCDL 2002), pages 171-180. ACM, 2002.

[10] C. J. Prom, C. A. Rishel, S. W. Schwartz, and K. J. Fox. A Unified Platform for Archival Description and Access. In Proc. 7th ACM/IEEE Joint Conf. on DL, (JCDL 2007), pages 157-166. ACM, 2007.

Abstract
Institutional repositories are currently evolving into a single core of digital collections. Necessary for this evolution is the task of charting the principles used in creating a matrix, which can be derived from the rules for the preservation of cultural heritage in order to endure over the course of time. These principles should be derived in such a way that they develop far beyond what the usual technology-dependent advice and methods would. Technologies come and go, but preservation is a task that should be carried out independently of the currently common line of thinking. Technology with know-how in preservation matters is only one aspect. Preservation is a cultural mission, which must guide our actions. The code that we select in the task of preservation, makes our culture bound behavior reproducible.
Keywords: Open Access; open university; information storage; dissemination of information; cultural policy

Introduction

In 1984, Italo Calvino was officially invited by Harvard University in Cambridge, Massachusetts, to present the "Norton Lectures" which appeared in the preface of his posthumously published book(1). These were a series of six lectures focussing on the topic of "Poetry". This referred to any form of poetic communication, namely, of any: literary, visual, musical form. This brings us to the heart of the matter, since the objects of his studies also refer to objects with a multimedia format that are stored in our current digital repositories for posterity. This is especially true for institutional repositories in the scientific establishment, which is now evolving toward a core of digital collections and whose job is the careful storage of cultural heritage, including all teaching and research output in such a way that it remains accessible, usable and understandable over the long term.

Italo Calvino, as is further stated in the preface, was almost obsessed with clarifying this topic. In this passage, he presented some literary values that needed to be preserved for the next millennium. Calvino named these values as: "Six memos for the next millennium." These are in order: lightness, quickness, exactitude, visibility, multiplicity and, not carried out by him, consistency. With that which Calvino meant in his Norton Lectures on poetry, the author of these lines identifies himself and thus, makes the following issues the focus of his essay:

- What principle is recognized in the field of preservation of cultural heritage and more specifically, how can one use it to create a code that is both derived from universal rules and can be reproduced at any time?
- What are the key ideas and possibilities available to us that will withstand deterioration?
- Technologies come and go. What would be beyond the usual technology-dependent advice and methods of use that would enable us to conduct the act of preservation effectively at all times, regardless of the respective current mindset and the technological platform?
- After these considerations, a momentous question should also be strongly posed: Why are the local research funders so quiet on this topic??

Closing the Generation Gap

The cultural differences between the younger generation and their parents are expressed today in such a way that we experience in the field of "new technologies in the workplace" the following remarkable fact: the parental generation, in the application of these technologies, obtains the support of their next generation and, consequently, often learns from their next generation while dealing with these technologies. The transfer of know-how and experience therefore occurs in persons from younger to older alike. At the same time, we are seeing for the first time in history, and within the last few years, how a rich, technical apparatus is created – accompanied by the respective know-how – for which the particular content literally has yet to be found. This is now being created mostly by commercial "content providers" with the benefit of hindsight. Formerly it was the reverse: first came the content (presentation of a problem), and then came

the solution as a derivative of know-how. In summary: Technology today is the "form", and chronologically speaking, is first in the world, whereas only until the second instance does one wonder, "cui prodest?" Content or various sources of content then result, because they were made for this purpose. Does technology now drive methods and processes or vice-versa? Now it behaves in such a way that today's administrators of their cultural heritage have been trained at a time when there were other methods and decision-making processes in the scientific establishment, and they therefore often applied an apparatus that can be derived from this earlier thinking. It is not timely. These circumstances lead to a strengthening of the generation gap.
Corollary to the formation of a matrix: The methods handed down in the scientific establishment from one generation to the next as well as decision-making processes should be called into question to the extent where the following principle may come to bear: What really matters is not whether a particular system is perfect or true ("true" is an arbitrary term), but rather how well it functions for the respective user/user groups. Efficiency is the measure of this "truth".

Digital Archives and Cultural Heritage: Rethinking Workflows, Administration, Decision-Making Processes

The last years were marked by the so-called convergence of technologies. What was left out of the general discourse was the convergence of knowledge (or better stated: the convergence of disciplines). The focus was not, as apparently widely believed, on extreme specialization, but on the convergence of disciplines. The projects that are practiced today arise from this convergence. Digital archives can be taken as an example here. Really successful projects are solely those projects where different disciplines are integrated into one: law, computer science, linguistics, psychology, philosophy of science, communication science, economics, sociology, and so on and so forth.
Admittedly, the Internet was created in academic circles and used for the first time on a broad basis. However, there are a large number of findings, knowledge-related processes and functions that our academic communities are only just now beginning to understand on a broader basis where they in turn then very slowly process such data. Chronologically this takes place after someone else, namely the next generation, has already successfully used it. In this context, the ease with which one can take advantage of the terms for digital repositories offered by Calvino for the Norton Lectures, is truly amazing.
First, a brief note on the term consistency: here, the "shelf life of the data" is meant and the remarks on preservation can be found again throughout the entire text.

Multiplicity: This not only stands for the consistency of the data (e.g. metadata) and of work processes including the complexity of the resulting processes in the digital archive (e.g. with preservation and functions of reuse), but also the complexity of the system itself. The next generation has been practicing networking and the building of communities for years. The young students of today are the young scientists of tomorrow. They are the ones who cavort in the social web for years and for example, create lists of "friends" on their Facebook pages and link to multimedia content. That is exactly what they are doing, deliberately and consistently, including the citability of dataset: How many repositories would have to be made available to our traditionally run, educational scientific institutions in Europe in order to benefit – assuming academic methods – similar communities of students and their teachers involved in research?
Corollary to the formation of a matrix: It is time to create similar codified opportunities for our repositories, such as chat rooms for setting up a digital repository. Another measure would be to allow in our universities access to the functions of the systems in order to open them up also to guests (befriended scientists or partners in a project). Privileges and access policies should be the same as for the "own community".

Lightness: Lightness requires doing without the exclusive use of centralized logic, central systems, and "central intelligence".
In the context of digital archives, the lightness of operation and the "lightness" of the data are essential characteristics one expects from digital archives. This lightness is synonymous with intuitive controls and should be realized by using common and generally accepted standards. A generally accepted standard does not necessarily mean "a certified standard". The "lightness of the information structured in accordance with generally accepted standards" may be the lowest common denominator of the demands of all concerned, qualified providers of data to a scientific digital archive. Defining "lowest common denominator" could result in the following task: The task of the system is the fostering of

communication between those things that are different. In this case, the differences would not be blurred, but, on the contrary, strengthened by highlighting the characteristics of individual digital objects. (Regarding the digital objects: no matter what format and type, they must be provided with contextual information and equipped with technical, descriptive, and long-term digital preservation metadata at their inception).
We have seen how some platforms have prevailed worldwide. What each of these have in common is namely their lightness. I will mention four of them at random: Napster (who can still remember the peer-to-peer exchange of data and the author alert systems used then?), eBay, Amazon, and YouTube. Let's stick with YouTube for a moment and take as an example the lightness of use in uploading content. This is not just about the operation, here the focus is primarily on access to the data delivery process (anyone can upload content) and on access to information (anyone can download content). The lightness is shown in a further example: The next generation uses Flickr as one of the world's best online photo management and sharing applications. Why doesn't something similar happen in the scientific establishment?
Corollary to the formation of a matrix: The scientific community should have the complete opportunity to store images, pictures and powerpoint presentations in institutional repositories of their institutions quickly, easily and inexpensively (see also multiplicity). Here the emphasis is on the processes of targeted publishing, quoting, commenting, sharing and reuse by and with the interested public, or community.
Another corollary to the formation of a matrix: The previous corollary implies that digital content is reliably archived long-term, provided with appropriate metadata, and more easily and always searchable via a persistent signature (assignable to the digital content and/or the respective author with a digital author identificator).

Quickness: The speed of the system is determined by the normalization of data (removal of redundancy) and by dividing it into areas that are associated with a specific task. The concept of the system operator should be determined primarily by efficiency and intuitive recognition.
The efficiency of the user interface – interaction concept – should be distinguished by the fact that the information content on the screen is not too compressed. From the above platforms, I now move to eBay. The operation of the platform requires a low level of literacy, although the services offered can be very complex (e.g. the clarification of payment terms, the resolution of legal issues and questions of logistics.) The user focuses only on his projects, all collateral duties will be met by the system (e.g. allocation of tags for indexing.) The same goes for Amazon. In all of these systems used worldwide, ethnic, political, or linguistic borders are irrelevant, rather, a variant of the game of accessibility becomes effective.
Corollary to the formation of a matrix: The assignability of information should take place quickly and in one effort. The user should always know where he happens to be, what he is doing and how he can cancel a transaction. He may never get lost in the system. Accessibility plays an important role. Accessibility is not just a purely technical issue to be solved, rather, it must be a fixed part of the deliberations at the stage of conceptual planning.

Exactitude: The accuracy of the descriptive data is to be achieved through standardization and coding. The submission of data should be conducted according to a set standard, supported by an information code for the individual entries. Thereby, one or more subject catalogs should be used, which are used to classify which have equivalents in other languages and which have cross-references. The systematic accuracy should be determined by syntactic accuracy (the syntax of the user input is determined and controlled through the system in each case), and by semantic accuracy. At this point, an incidental remark made in 1984 by Calvino on "exactitude" can be quoted. For him, exactitude was three things, and for this essay, I will employ namely his second point mentioned: "The evocation of clear, distinctive and memorable visual images, in Italian we have an adjective for this that does not exist in English nor German: icastico from the Greek eikastikós."(2)
Here it is quite remarkable how, still in subsequent years with the next generation, the term "icon" could prevail in everyday language in its wide-ranging application, and now in both languages precisely this semantic aspect in English and German has become indispensable.
Corollary to the formation of a matrix: Also in this case, the effectiveness is the measure of true accuracy. In general, no "blank spaces" should exist (i.e. "null values" for information derived from queries should not be allowed.)

Visibility: In this context, visibility is considered the ease of use. The user should be able to carry out actions based on on-screen information and interact with the system, or alternatively, to retain on-screen information as a consequence of actions the user takes. Furthermore, the degree of traceability is not only significant for the objects (e.g. the origin) but also for the work processes (that is, the traceability in the search history and its results).
Corollary to the formation of a matrix: Not only the visibility of the repositories should be increased. Institutional blogs should be conducted at the interfaces to the repositories of our institutions. The digital content, the content of certified repositories, should be organized in relation to each other or linked (e.g. in order to create new collections of digital objects and to scientifically annotate them). Qualified content should be posted, linked and annotated. Even certainly the linking of the repository and long-preserved material should be linkable to and from platforms like, for example, Twitter. We need a cross-disciplinary approach.

Limits to the Digital Preservation of Cultural Heritage

What prevents us from following these corollaries? Mainly honored traditions and processes (methods), even if some or even all of the above conditions are satisfied. In addition, crucial is the lack of confidence in the know-how developed to date, as well as in the expertise of active promoters of these processes. The next generation (and a part of it is our young scientists) is often unfortunately not taken seriously because they are considered too young (which is also the reason why this writing does not include something about Twitter-like services nor something about the possibilities that would result from safeguarding cultural heritage for use in mobile applications. At this very instance, developed know-how in general seems methodologically too young.
Corollary to the formation of a matrix: We need to foster the building of confidence in expertise, competence and the available e-infrastructure. It would also be recommended to develop certification mechanisms for digital archives, so that reliability would be resultant and thus the citability of datasets would be enabled. Quality Assurance in all of its facets would then be a part of the certification mechanism.

Conclusions

In safeguarding cultural heritage, we need a more sophisticated way of thinking for developing solutions. The approach to the design of systems – including systems of thought – should be crossdisciplinary. We need a crossdisciplinary approach. The instrument itself, the digital repository, should be designed from the beginning as a multimedia marketing tool that enables information transfer and communication between users (data suppliers and consumers) in order to make content consistently available. The possibilities for distribution of information would be far more diverse and knowledge would not only be stored but its implementation would be easier. To accomplish this, a different access system would have to be designed from the beginning, including a sophisticated rights management system. Of course, the data provider should retain all sovereignty over the data. These are not empty words as the solution for these issues today is not of a technical but exclusively of a political nature. Access and restrictions (technically and legally speaking) are mostly an expression of political will. The same goes for accessibility: it is not only challenged technically, but also in all of its associated processes, in data production ranging from the delivery of data to the final data output.
We should therefore redefine the role of users. Users can, in principle, be individual users or institutions. Users can also be subdivided into groups of data providers and consumers. This necessitates a policy of open and free access not only for the consumption of published information, but also in the very process of publishing itself.
With regard to users, generally speaking, they should be empowered, especially that user access should be enhanced, particularly in the following two roles and processes:
The user as a data supplier with free access to the digital archive
The end-user as the beneficiary of the digital content in the digital archive.
It should be possible to guarantee this end-user free access to all published information. He should be given more rights and functionality. This requires a different approach with the wishes of the end users (focus groups) on the system. For example, for the purposes of the reuse of digital content, the end-user should be in a position to be able to implement the new knowledge gained by linking it with other content online (e.g. formation of collections of data sets inside of the repository). In addition, he could be empowered to link individual objects together so that he can therefore "form virtual dossiers" which he can them make available to other users in his community.

Interoperability with other systems and trust in online interaction should be guaranteed to the user (individual user or institution) as a data supplier. For information providers, who are not from the next generation, special training should take place offering these users more expertise, especially in the areas of preservation and reuse of information and techniques of self archiving.
Finally, perhaps the most important corollary on the formation of a matrix comes this time in a personal form (and please forgive the repetition):
What really matters is not whether a particular system conforms to a "true" norm ("true" is an arbitrary term), but rather how well it functions for you the user. Effectiveness is the measure of this "truth".

REFERENCES:

[1] Italo Calvino, Lezioni americane. Sei proposte per il prossimo millennio.Garzanti, Milano, 1988.
[2] Cited from the German translation of Calvino's work, page 83, in: Italo Calvino, Sechs Vorschläge für das Nächste Jahrtausend, Harvard Vorlesungen, Carl Hanser Verlag, Munich Vienna, 1988

Abstract
Social Media is transforming visitors' behavior in Internet, they become active participants in the creation of knowledge instead of passive viewers. Memory institutions are slowly introducing those media to respond to the latest social demands, which involved opening their institutions to public contributions and opinions.
Public create content, develop local or distant virtual communities, through which interests and information are shared. People that belong to those communities express preferences, feel accepted by other peers and develop a more direct relationship with museum staff. Despite the lack of enthusiasm from curators and unsolved problems, the Web 2.0 phenomenon seems to offer new strategies to attract audiences and encourage users to get involved with their cultural institutions. The present paper attempts to discuss these new strategies by analyzing the aims and expectations as well as their comprehensive fears. Two different case studies in which we are currently involved, allow us to discuss in detail two successful experiences: the Immigration History Museum of Catalonia (http://www.mhic.net - Spain) and the Civic Museum of Rovereto (http://www.museocivico.rovereto.tn.it - Italy). Although both case studies present differences in the type of collection and objectives, they have employed the new social media to reinforce the existent real communities around both institutions. Keeping strong ties with their local communities also through Internet provide them with original content that can be attractive to distant visitors as well.
Our paper pretends therefore to develop a further reflection on visitors' cultural content creation and strategies that small and medium-size institutions with limited-budget can take on to disseminate their original cultural heritage.
Keywords: communities, cultural content creation, Web 2.0

Introduction
The "mediamorphosis" (Fidler, 1997)[1] currently taking place in cultural institutions has evolved as a tool to support the creation of relevant information. Museums already use Internet as a new mean of communication that allows them to access new publics. However, the development of social networks and the Web 2.0 supposes them a new challenge with potential huge benefits in terms of social response. Providing participatory tools is a way to satisfy public's requirements, who wish to express their own preferences and points of view to Museum curators. Although such freedom of speech may generate uncomfortable situations for those cultural institutions, the phenomenon represents new life for institutions with a virtual community around them.
Indeed virtual communities seem to be useful to attract new audiences because through social groups, people tend to establish emotional relationships (Rheingold, 1994)[2]. It is easy to imagine that through those Web 2.0 applications, lasting relationships can sprout amongst museum's users as well as between them and the institution itself. Museums ply new Web 2.0 tools to encourage virtual audience to repeat a physical visit to the centre, promote daily discussions and often involve visitors with empathy to become "friends of the museum" (Von Appen, Kennedy, Spadaccini, 2006)[3]. Besides, creating virtual communities increases the average time spent in the website, so a feeling of belonging towards the institution (Sommavilla, 2007)[4].

[1] Fidler, R. (1997). Mediamorphosis, Understanding New Media. Colorado, Pine Forge Press.
[2] Rheingold, H. (1994). The Virtual Community: Homesteading on the Electronic Frontier. Secker & Warburg. London. Online (updated) edition available from: http://www.rheingold.com/vc/book/ [Accessed 23 September 2005].
[3] Von Appen, K., Kennedy, B. and Spadaccini, J. (2006). Community Sites & Emerging Sociable Technologies, in J. Trant and D. Bearman (eds.). Museums and the Web 2006: Proceedings, Toronto: Archives & Museum Informatics, published March 1, 2006 at http://www.archimuse.com/mw2006/papers/vonappen/vonappen.html editor's note. URL corrected Jan. 21, 2007.
[4] Sommavilla D (2007). Nielsen//Netratings comunica la dimensione del Web 2.0. Per la prima volta in italia uno studio quantitativo del fenomeno. http://www.netratings.com/downloads/nielsen_netratings_Web20.pdf.

Chiara Cirinnà and Maurizio Lunghi (edited by), *Cultural Heritage on line: Empowering users: an active role for user communities. Florence, 15th - 16th December 2009* ISBN 978-88-6453-184-7 (print) ISBN 978-88-6453-187-8 (online)

Likewise, the movement of cultural information, as well as producing very effective phenomena of viral marketing, encourages interested parties to select information according to personal preference and stimulates production of content and users' commentaries. Such content gives fresh-air to traditional institutions such as museums since it is continually renewed by the own public. This democratic approach becomes more reliable to the eyes of other users and at no cost for museums, whose only task would be limited to check the correctness of the received material.
Despite these undeniable advantages, it is necessary to combine appropriate strategies with the use of social networks. Otherwise, social media are likely to remain sterile tools, sometimes even harmful to the own aims of cultural institutions. Indeed there are a lot of unfortunate examples of museums that end for abusing the advertisement mailing without drawing a suitable communication strategy. Similarly many of the tools used are often unable to stimulate the creation of communities around the museum, leaving many opportunities offered by technology, unfathomable.
As use of these tools is a fairly recent phenomenon, it is appropriate to do same research on how they can become effective in institutions' strategies. Cultural institutions reflect their own identity in the way they are seen by public. Therefore, communication strategies are key issues in the future museum projection outwards, becoming a kind of institutional DNA. Therefore, the advent of communications and information technology, gradually introduced by institutions for enhancement of cultural heritage, should be used according to the institutional aims, celebrating its distinctive characteristics and peculiarities. An unconscious use of them increases the risk of deforming the identity of the institution in favor of momentary trends, perhaps destined to disappear in a few years. For this reason, here we aim to analyze two cases with which we were lucky enough to work directly. Both, while adopting two different approaches towards the use of participatory technologies, seek to combine the contribution of users in the production of knowledge with the original mission of the institution, representing two success experiences on which to center our reflection.

The museum as a services' lab: the case of the Museo Civico di Rovereto

(http://www.museocivico.rovereto.tn.it)
The Civic Museum, since its opening in 1855, has been defined an institution closely related to its territory. Its main purpose is recording the environment where the community lives, in which citizens can rediscover their own roots, obtain local information and interact with the social and economic tissue. Following such innovative tradition, the Civic Museum of Rovereto decided to open the institution to technological developments and new opportunities' offered by Web 2.0 while remaining faithful to its objectives. The combination of tradition and innovation leads in the case of the Civic Museum to provide services to its citizens and to a community of local actors willing to invest the needed resources for sustaining the museum. The collection of scientific data obtained by its historical equipment, combined with an ongoing dialogue with the productive forces for the promotion of knowledge and citizens' participation, have therefore stimulated a community cooperation in which the museum is rooted.
The website of the Civic MUseum of Rovereto (http://www.museocivico.rovereto.tn.it), for example, offers hundreds of thousands of digitalized cards related to the different fields, providing the public with an archive, continually updated, also available for online access. Some of these cards are geo-referenced, with a GIS supported by the University of Siena, and regularly consulted by experts and professionals of different sectors after subscription.
The virtual space shows both the artifacts owned by the museum as well as an interconnected network of additional information which informs citizens about a phenomenon that occurs in its territory educating them to the use of precautionary measures. External collaborators of the Civic Museum, specially trained for updating the portal content, exchange and transform data gathered in a collaborative way in the fieldwork and made it available to the public in real time.
The Web Directory is reserved to research groups to share scientific tools and information. Access to that space is made possible just by requesting it to the museum. Therefore, the Museum has developed a kind of virtual lab for an active scientific community related to the local museum, which also disseminates such specialized information very fast.
The museum, as a great cultural workshop, can manufacture and sell the knowledge which is constantly added to its container thanks to the community cooperation. Some environmental phenomena taking place around Rovereto, for example, were monitored through citizens' contribution and scientists who scanned and shared the results of their investigation on the web page or on the Web Directory. This approach gives identity to the community and encourages his actors to support the museum by participating in its activities. Especially the latent communities and the ones active just in a physical way are being reinforced by the museum's intervention and by the real and virtual initiatives launched by it.

The dialogue between the productive forces and the great work of economic animation has created enormous repercussions not only in terms of money but also for contacts and new opportunities. The budget of the Civic museum for example is due only in small part to the revenues of public institutions. The rest of the cost of the structure is covered by selling services to businesses, professionals, governments and to a variety of subjects demonstrating the effectiveness of the model undertaken.
Another important information channel through which the museum proposes scientific news in video format is Sperimentarea.tv (http://www.sperimentarea.tv/). The scientific Web TV is an open laboratory, which combines museum's experiences with students' creativity, teachers, researchers and professionals. The television station on air offers a diverse programming broadcast on a fixed schedule besides sections of video on demand where users can choose both large movie productions by international filmmakers and curious, interesting, explanatory movies uploaded by researchers, students and users. Contents can be viewed on any PC or portable audio-visual media, including mobile phones and sent to other users with a function of email alert. Using the audio-visual medium, therefore, the museum draws the attention of the general public, showing how science and technology disciplines are at the service of people. Thanks to these initiatives provided to it citizens, the Museum has managed to cluster around them a vast network of experts, local businesses and citizens, strongly motivated to cooperate with the institution because of its services strongly useful to all of the residents in the area. These networks of contacts allow the museum to be self-sustainable. Choosing to sell services it opted for the development of a virtual environment accessible only to the professional community, to ensure the quality and reliability of the published information. Analyzing the log data from the Museum website, it is remarkable that users of the Civic Museum download 50 gigabytes per year on average. Although the time's visit is long, it is assumed that users accessing to the platform are really interested in content posted or services provided. The contact established with some of them through the museum in fact confirmed to us that many of them were professionals and experts, who access the virtual resources of the Civic Museum often because it offers information not available through other sources.

Life stories': the Immigration History Museum of Catalonia

(http://www.mhic.net)
A completely different approach is offered by the new-born Museum of Immigration History of Catalonia, which opened in 2004 in Sant Adrià de Besós, a small-town in the neighbourhood of Barcelona (Spain). The Museum attempts to records the life experiences of all the immigrant people that came to Catalonia and the city of Barcelona for a job opportunity and finally settled down here. As happens in other regions and countries all over the world (i.e. USA, Australia, Argentina…), their progress and wealth was due to some extent to the labour force that came from other regions. These anonymous stories are not normally part of the traditional history Museum that is why Immigration Museums have grown as independent institutions.
Sant Adrià de Besós is an immigrant town, whose people came from different parts of Spain and nowadays from other countries (i.e. Africa, America and Asia) to work in the local industries and construction. The initial aim of the Museum was to create special links with the local community, because the real collection of the Museum was the testimonies of anonymous people who wanted to explain their story.
At the time the Museum was being developed, the UOC (Universitat Oberta de Catalunya) was involved in a European project called COINE to generate a digital archive on-line with testimonies of local communities. It was believed that the Immigration Museum was an excellent field for a pilot experience, so the application was implemented here at the time the portal was being created.
It was such as sophisticated application with metadata tagging and thesauri, as well as multimedia files (i.e. audio, video, images) that most potential content providers, who were old people, were quite afraid of taking part. Therefore, different tests of usability demonstrated that complex applications could not be attractive enough for old people, who did not want to invest much time in explaining their life story.
On the contrary, they could spend sometime in front of a videocamera or taperecorder, so museum curators recorded their stories in those formats. When the social media started to become popular, all those testimonies were posted in channels such as Youtube for video or Odeo for audio. Despite the fact that this was a possible solution, it involves people coming to the Museum and arranging a date for recording themselves. Therefore, the advantages of Internet as a way to break barriers of time and space could not apply here.

Old people were afraid of recording themselves on-line because they require an assistant to help them in the first steps to play with multimedia formats as well as the sophisticated metadata tagging, whereas a presential alternative was too time consuming.
Social media, in this case multiuser-blogs, have become an excellent alternative for the Immigration Museum. The idea behind was a collaboration with the local schools asking students to become assistants of their families relatives. Students for particular schools and courses under the supervision of their teachers have been providing life's stories of their relatives including any kind of media (i.e. images, text, video). They write the story, scan images or provide digital ones and videos if they have and publish in the blog.
Here, the most important issue is who controls the quality of the life's story, in this case the own teacher. Schools take part in the activity as part of their own curricula and an interesting way to show concepts such as multiculturality and identity. Young students from different origins explain immigration stories that have common traits with their own schoolmates. The possible fears of the Museum of publishing inappropriate contents are vanished since contents are controlled by teachers before publishing. As you can see, it is not opened application of social media, but a regulated one based in an existing local community that provide contents to the institution.
The final aim of the Immigration museum is creating a completely digital archive that combines objects, documents and memories from the own museum collection obtained from researchers linked to the institution together with contributions of the local community. However, the more data is updated in this virtual archive more need will be to create somekind of metatags to favour intelligent search in such database. Probably, this documentation task should rely on the Museum staff. Combination of specialised documentation with local contributions make the Immigration Museum archive quite an interesting website for experts, which have widely used so far those documents available. One of the problems that should be address is how to allow other users non-related to the local schools to introduce their own stories. Probably, another Web 2.0 application could be the answer to such requirement, but there are still questions about administration and content control that ought to be born in mind.

Conclusions

Thanks to these two cases analyzed, it is possible to see how the public input in the production of content represents a huge resource for cultural institutions. Through the introduction of participatory applications it is possible to organize a network of contacts that contribute to raise the importance of the web page and as a result of the museum. However, as mentioned previously, it is important that the participative applications used are accompanied by strategies that involve users, motivating them to become concerned. To do so the museum should promote the creation of communities around it and prove useful services to the public.
The web services for example can be structured primarily on the needs of the local community as well as on the objectives of the institution. The resources that the local community can offer are enormous if coordinated with the services and benefits offered by the museum. These if properly designed and accessible to the public can motivate a large number of users to provide their contribution. To offer new services to citizens or to the community around the museum it is necessary to maximize the potential of existing technologies and find the appropriate tools and strategies to achieve the museum's objectives. This is supposed to develop a participative approach in the creation, use and administration of local cultural content that meet practical needs of information and learning.
In fact the activity of a museum (organization of exhibitions and events, renovations, acquisitions, etc.) brings the institution to collaborate with different communities and offline groups: students, scientific associations, schools and voluntary associations. It is also important to encourage this "public" to join the museum's online conversation, providing them specific content and tailored web spaces. Likewise the on-line activity of the museum should create the conditions for groups of people with common interests to join and form an active virtual community on the web.
If there are functional needs (Giacoma, Casali, 2008)[5] in the local tissue to which the museum is able to answer it is very likely that the community can grow. The museum should also be able to design a strategic network to meet practi-

[5] Giacoma, G., & Casali, D. (2008). Elementi teorici per la progettazione dei Social Network. Tratto da Issuu: http://issuu.com/folletto/docs/elementi_teorici_per_la_progettazione_dei_social_n

cal users' requirements and at the same time to please those relational motivations (Giacoma, Casali, 2008) triggered in social contexts, the satisfaction of which leads to the recurrence of the experience. To enhance users' motivation to participate it is therefore necessary, besides offering specific services, to stimulate their curiosity as well as their desire to share interests and to feel part of a group. 2.0 applications that currently are spreading in the portals of museum's institutions therefore require:

- An active community considering them a means for obtaining some benefits whether they are informative, fun, learning or otherwise.
- Strategies for monitoring of content produced by its community.

Institutions have two options using social networks open to users' contributions. The first one is to have a team of professionals responsible for the control of the material uploaded by users and for the verification of its ethics and honesty. The second one is to leave the community free to regulate and manage inappropriate content as in the Wikipedia. Unfortunately, this second option works only if the active community is composed by a large number of users that can adequately take care of removing or correcting inappropriate or offensive comments. If not there would be the risk of leaving unsuitable material online for a long time or fail to refute or correct the improper content before others use them. This could greatly lower the quality of services offered to the community and undo the benefits that a cultural institution should be able to ensure to its audience.

Small and medium-sized institutions cannot always afford to devote internal resources to the review of the content posted by users because of budget and at the same time when they decide to use participatory applications they do not yet have virtual communities around them able to supervise the content in an independent way. The participation of communities already physically active in the museum, as well as the latent ones, can therefore be fundamental to support small institutions in managing users' contributions.

In both the analyzed cases for example it is guaranteed the quality of published users' content through two different strategies derived from different stories and objectives, but equally effective because based on an active local community, interested in the outcome of its collective collaboration. Virtual communities, born from a real local need can increase and include new distant communities by offering interesting content and a strategic example of how sharing life experiences and knowledge. These two cases thus represent two examples of good practices from which to take the cue when open the door to the plurality of voices of the network.

Abstract
This paper briefly deals with a digital application designed for presenting the Roman Agora and the Library of Hadrian, two adjoining civic structures situated in the historic center of Athens. Although they have suffered destruction and alteration in form and function over the centuries, they have played an active role over the years in the life of the city and are at present two major archaeological sites of Athens.
The paper is divided into three parts. The first involves the architecture and history of the buildings. The second part concerns the scope of the project, and the final section gives a short description of the digital application.
The project is conceived as a virtual tour through the two monuments allowing the users to explore them interactively feature by feature and phase by phase. The application includes maps, plans and perspective reconstructions of the monuments, engravings by travellers, photographs, QVR Panoramas, Google Maps and informative texts. A dynamic timeline allows users to follow the most important historical events concerning the city of Athens during the Roman, Byzantine, Ottoman and Modern era. Personalities associated with the history of the monuments are highlighted and specific architectural terms elucidated.
The on-line digital application was designed and financed by the Directorate of Museums, Exhibitions and Educational Programmes of the Hellenic Ministry of Culture and was produced and animated by the company Minimatik, visual + interactive communication and coordinated by Makebelieve, design & consulting. As soon as the project is completed in both Greek and English, free access will be possible through the official website of the Hellenic Ministry of Culture (www.culture.gr). One of the goals of this project is to function as a model in the future for similar applications dealing with other monuments all over Greece.
Keywords: The Roman Agora, Hadrian's Library, virtual tour, Athens, archaeological sites

Introduction
The aim of this paper is to present briefly an on-line digital application for the Roman Agora (=Market) of Julius Caesar and Augustus and the Library erected by the Roman emperor Hadrian at the heart of Athens, the metropolis of Classical civilization. These are monumental building complexes, initially two porticoed enclosures, lie next to each other in the historical center of the capital of modern Greece (Hellas) and, albeit altered in form and function, constitute a major landmark of its topography.
This on-line digital application is a virtual tour through these two monuments. It was designed and financed by the Directorate of Museums, Exhibitions and Educational Programmes of the Hellenic Ministry of Culture and was produced and animated by the Greek company Minimatik, visual + interactive communication and coordinated by Makebelieve, design & consulting. As soon as the project is completed in both Greek and English, free access will be possible through the official website of the Hellenic Ministry of Culture (www.culture.gr).
These monuments are ideal for a digital presentation in view of:

- their elaborate architectural form and function,
- their long history,
- their nodal location within the historical city center and
- their modern function as major and popular archaeological sites.

The state of preservation of both structures varies from fair to poor and the ruins seen today represent a mixture of different building phases. As a result, understanding the Agora and the Library is a challenging task both for tourists and scholars. This bilingual digital application presents them concisely and accessibly to the public worldwide.
The present paper is divided into three parts. The first deals with the architectural type and history of the buildings, the second with the objectives of the digital application and the final section offers a short description of the project.

Chiara Cirinnà and Maurizio Lunghi (edited by), *Cultural Heritage on line: Empowering users: an active role for user communities. Florence, 15th - 16th December 2009* ISBN 978-88-6453-184-7 (print) ISBN 978-88-6453-187-8 (online)

The Monuments: brief presentation

Figure 1 - The Gate of Athena Archegetis

The Roman Agora and the Hadrian's Library were built in Athens under Roman authority. The former was financed by Julius Caesar (51-47 BC) and Augustus (19-11/10 BC), whilst the latter was conceived and donated by the philhellene emperor Hadrian (131-132 AD). Their construction reveals the personal interest of Roman rulers in Athens.
The Roman Agora

The Roman Agora (1) was built on the site of a crowded open market, which extended up to the principal commercial and civic center of Classical Athens, the famous Athenian Agora. The Roman Agora consists of a rectangular building complex around an open court surrounded by porticoes (stoas). Its plan, which is a quadriporticus in form, recalls that of Roman fora, which catered for religious, political, military and commercial activities. Its two entrances, which face each other, are enhanced by monumental façades (propyla). At the west gates, known as the 'Gate of Athena Archegetis' (=the patron goddess Athena), the road connecting the Ancient Agora with the Roman Market terminated. This road, paved and flanked by shops, was reserved for pedestrians and was called the Wide Road. The Ionic façade on the east side was placed off-center, since it marked the end of an old street. The north side of the structure has not been completely excavated. The Roman Market at Athens had storerooms (horreae) on the west side, shops (tabernae) on the east side and a fountain on the south side. Although some offices connected with the operation of the market (e.g. control of prices and weights) may have been housed in rooms across the south side of the structure, it would seem that the office of the market officials is to be sought west of the Roman Market. The paucity of shops may suggest that they were intended only for wholesale traders, whereas retail trade may have taken place in the court and the stoas.
Before the construction of the Roman Market a marble octagonal tower existed directly east of it (2). It combined a hydraulic clock in its interior and a sun-dial and vane on its exterior. This sophisticated ancient mechanism, an invention of Andronicus of Cyrrus, in Macedonia, is known as the 'Clock of Andronicus from Cyrrus' or 'Tower of Winds', due to the relief frieze around the upper part of the exterior of the building that displays personifications of eight winds. This unique structure is apparently a creation of the 2nd century BC, a period when technology excelled and developed rapidly.

The Roman Agora from the Byzantine to the Modern period

Figure 2: The Church of Prophet Elijah & the Church of the Taxiarchis, Engraving by Th. du Moncel.

A three-aisled basilica was built within the Roman Agora during the Early Christian period (4th-6th century AD). It was converted into a mosque in the Ottoman period (1456-1830 AD). Likewise, in the Early Christian times the Tower of Winds served as a baptistery of a nearby church, whilst in the 18th century it housed an Ottoman Tekkés (holy place). Two adjacent domed cross-in-square churches, the Church of the Taxiarchis (3) and the Church of Prophet Elijah (4), were erected on the north side of the Market complex in the Middle Byzantine period (11th-12th century AD). The Church of the Taxiarchis was demolished in 1852 and was replaced by a new church dedicated to the Archangels and the Virgin. The Church of Prophet Elijah was refurbished in a rudimentary fashion after the Greek War of Independence (1821-1828 AD), in order to serve as a hospital and in 1848 it was finally demolished.
During Ottoman rule in Greece the little Church of the Soteira tes Pazaroportas, dedicated to the Virgin, was built on the north end of the West Gate of the Market. It was demolished after the liberation of Greece (1829 AD).

Hadrian's Library

Figure 3: Hadrian's Library

The Library (5), directly north of the Roman Market, was a rectangular two-storied building around a peristyle courtyard with a pool in the middle. It was provided with reading rooms, spaces for the storage of papyri and with lecture halls. The building was approached from the west through a propylon with four Corinthian columns of Phrygian marble which was flanked on both sides by a row of seven columns located on pedestals of green Karystian marble. This substantial intellectual and cultural center of Athens, which also housed the archives of the city, was badly damaged during the invasion of the Heruli, a Germanic tribe, in 267 AD.

Hadrian's Library from the Byzantine to the Modern period

In the early 5th century AD, the Library of Hadrian was repaired. During the first half of this century, the central pool was filled in and on this spot a luxurious tetraconch church (6) was built by the Eparch of Illyricum, Herculius, or, in the view of other scholars, the empress Eudocia, a native of Athens. At the end of the 6th century, the so-called Tetraconch was severely damaged, probably because of some Slavic invasion, and converted into a three-aisled basilica, which was destroyed at the late 11th century AD.

A little, single-aisled domed church, the Megale Panagia (7), dedicated to the Virgin, replaced the basilica. At its south end a second church, dedicated to the Holy Trinity and of the same type and size, was annexed during the 17th century or a little after 1715. After the liberation of Greece (1828), the Megale Panagia housed the state collection of antiquities. It was demolished, after being burnt, in 1885 to allow archaeological excavations. The Hagioi Asomatoi "sta skalia" (8), a church dedicated to the Archangels, abutted the north colonnade of the west monumental façade of the Library during the 11th-12th century AD. From 1576, when the church was renovated and decorated with wall-paintings, if not from the time of its construction, it belonged to the eminent Byzantine family of the Chalkokondylai. This church in its turn was demolished in 1849.

During the Ottoman occupation of Athens, a commercial center with more than 100 shops, known as the "Upper Bazaar", grew up in the area of the Library. This extremely lively area operated until 1884, when it was destroyed in an enormous conflagration. In the southwest corner of the Library the residence of the Turkish governor of the city, Voevodaliki (9), was also erected. In 1835 the governor's mansion was converted into barracks and later into a prison.

The Roman Agora and Hadrian's Library at present – the scope of the project

Today the Roman Agora and the Library of Hadrian form two adjacent archaeological sites, where systematic excavation, restoration and rehabilitation continue. Their location at the heart of the historical center of Athens, in the pivotal area of Monastiraki, opposite the Metro station of the same name, makes them a familiar landmark for Athenians and a popular sight for both Greek and foreign visitors. Very few, however, are aware of the historical events associated with their erection, the politics pursued through it, their architectural prototypes, their various building phases, and the drastic changes in their use over the twenty centuries of their existence.

The digital application informs its visitors of these matters by means of a virtual tour through the monuments, in time and space. References to related events and personalities, a combination of texts and images presented interactively make the e-wandering an engaging experience.

The goals of this project are:

- to use the Internet as a medium to permit free use and easy viewing of the monuments by people all over the world,
- to invite the visitor to the website, whether an ordinary inhabitant of Athens who is in the habit of hurrying past the monuments or a potential tourist, to stroll through them,
- to help users grasp the history and cultural context of the monuments and their individual features and recall this information when they finally visit the sites in person,

- to offer a better understanding of the monuments for those who have already visited them and
- to function as a model in the future for similar applications dealing with other monuments, which can then be produced in co-operation with various provincial Ephorates of Antiquities of the Hellenic Ministry of Culture.

On-line digital application for the Roman Agora and Hadrian's Library

The aim of the on-line digital application for the Roman Agora and Hadrian's Library is to supply students, scholars and the general public with easily accessible, up-to-date and expert material on a digital and visually dynamic platform. The concept behind the navigation design is to invite users to experience the two monuments, rather than simply to access information, by taking advantage of the interaction possibilities the web offers as a medium. As part of this goal, a virtual tour and dynamic timeline were implemented, thus allowing users to explore the two monuments feature by feature and phase by phase and to follow the complex patterns of construction, modification and destruction from antiquity to the present day. Thematic essays written and reviewed by experts are accompanied by a plethora of images, plans and maps and two QVR Panoramas. A simplified version of the virtual tour has also been made available in a customized fully functional Google Maps environment, which offers a completely interactive experience.

Virtual Tour & Timeline

A two-dimensional interactive resizable plan including nearly thirty clickable individual features, including buildings, parts of buildings and major monuments, inside and near the two monuments, functions in conjunction with an interactive timeline. The visitor can navigate the monuments interchangeably, by place and/or time.

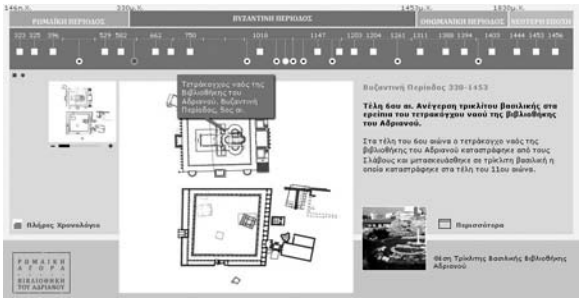


Figure 4: Snapshot detail of the Virtual Tour and Timeline.

The timeline is divided in four main periods, Roman, Byzantine, Ottoman and Modern. Clicking on one of the four periods highlights all the monuments that were either built or reconstructed during this period. Each period is subdivided in a number of dates. Hovering over a date displays monument-specific and general information, thus helping the user grasp the history and cultural context. Clicking on a specific date highlights the respective feature in the plan and brings up basic information, in the form of text and image, regarding the feature, within the current interface. Clicking on an individual feature in the plan highlights the feature and corresponding date, while realigning the timeline if necessary, and bringing up basic information.

The user may view complete information on the feature by clicking on a 'read more' link which opens a minisite in a new window. The minisite contains extensive information divided by tabs (history, description, special features, references etc.). Each tab contains an image gallery, corresponding to the information provided.

Main Navigation – Information Indexing

The website offers a different means of accessing the rich information associated with the monuments, dividing it into thematic sections (Monuments, People, and History). This twofold method of accessing information caters for accessibility issues and ease of use. Thus the user may choose how to navigate, either through lists or engaging in the interactive experience of the Virtual Tour & Timeline section.

Google maps & QVR Panoramas

Another section of the website is focused on the present. The monuments are localized in their contemporary environment through the use of the Google maps application and overlays. Two QVR Panoramas provide a current view by means of video footage, accompanied by texts offering an ideal walkthrough.

CONCLUSIONS

The combination of archaeological data and modern technology is now a reality and one of the objectives of the Hellenic Ministry of Culture. It is desirable that users of Internet be acquainted with Greek monuments by electronic means before their physical visit to Greece. The digital application presented above is the first effort in this direction, but it is hoped that it will serve as a model for similar projects pertaining to other monuments all over Greece.

REFERENCES

[1] J. M. Camp, The archaeology of Athens, New Haven-London 2001, passim. M. Hoff, The Roman Agora at Athens (diss. Boston University 1988). Idem, "Early history of the Roman Agora at Athens" in: The Greek Renaissance in the Roman empire, Papers from the Xth British Museum Classical Colloquium, BICS: Suppl. 1989, 1-8. Idem, "The so-called Agoranomion and the Imperial cult in Julio-Claudian Athens", Archäologischer Anzeiger 109 (1994) 93-117. T.L. Shear, Jr., "Athens: From city-state to provincial town", Hesperia 50 (1981) 356-377. A. Choremi-Spetsieri, "Πολεοδομική εξέλιξη και μνημειώδη κτήρια στην Αθήνα κατά την εποχή του Αυγούστου και του Αδριανού", Αθήναι. Από την κλασική εποχή έως σήμερα (5ος αι. π.Χ. – 2000 μ.Χ.), Athens 2000, 166-193. D. Sourlas, "Νεότερα στοιχεία για τη Ρωμαϊκή Αγορά της Αθήνας", in: S, Vlizos (ed.), Athens During the Roman Period. Recent Discoveries, New Evidence [Mouseio Benaki, 4th Supplement], Athens 2008, 99-114.

[2] H. J. Kienast, Ο Πύργος των Ανέμων. Οι Αέρηδες, Athens 2007. J. Von Frieden, ΟΙΚΙΑ ΚΥΡΡΗΣΤΟΥ – Studien zum sogenannten Turm der Winde in Athen, Rome 1983.

[3] Ch. Bouras, "The Middle-Byzantine Athenian Church of the Taxiarchis near the Roman Agora", in: J. Herrin, M. Mulett and C. Otten-Froux (eds.), Mosaic. Festschrift for A. H. S. Megaw [British School at Athens, Studies 8], Great Britain 2001, 69-74.

[4] S. Sinos, "Die sogenannte Kirche des Hagios Elias zu Athen", Byzantinische Zeitschrift 64 (1971), 351-361.

[5] J. Travlos, Pictorial Dictionary of Ancient Athens, London 1971, 244-252. D. Willers, Hadrians Panhellenisches Programm: Archäologische Beiträge zur Neugestaltung Athens durch Hadrian, Basel 1990, 14-21. J. Knithakis, E. Symboulidou, "Νέα στοιχεία δια την Βιβλιοθήκη του Αδριανού", Αρχαιολογικόν Δελτίον 24 (1969), Μελέται, 107-117. I. Tigginaga, "Η μεγάλη ανατολική αίθουσα της βιβλιοθήκης του Αδριανού (βιβλιοστάσιο). Αρχιτεκτονική μελέτη – πρόταση συντήρησης και αποκατάστασης", Αρχαιολογικόν Δελτίον 54 (1999), Μελέται, 285-326. A. Choremi-Spetsieri, "Library of Hadrian at Athens. Recent Finds", Ostraka 5 (1995) 137-147. A. Choremi-Spetsieri, I. Tigginaga, "Η Βιβλιοθήκη του Αδριανού στην Αθήνα. Τα ανασκαφικά δεδομένα", in: S. Vlizos, op. cit. (n. 1), 115-131.

[6] I. Travlos, "Το τετράκογχο οικοδόμημα της Βιβλιοθήκης του Αδριανού", Φίλια έτη εις Γ. Ε. Μυλωνάν, I, Athens 1986, 343-347. Idem, Πολεοδομική εξέλιξις των Αθηνών, Athens 2005 (3rd ed.) 139, 141.

[7] Ch. Bouras, "Επανεξέταση της Μεγάλης Παναγίας Αθηνών", Δελτίον της Χριστιανικής Αρχαιολογικής Εταιρείας ΚΖ΄ (2006) 25-34.

[8] E. Touloupa, "Ο Άγιος Ασώματος στα σκαλιά", Ευφρόσυνον. Αφιέρωμα στον Μανόλη Χατζηδάκη, II, Athens 1992, 593-600.

[9] J. Knithakis, F. Mallouchou, G. Tigginaga, "Το Βοεβοδαλίκι της Αθήνας", in: Ch. Bouras (ed.), Επώνυμα αρχοντικά των χρόνων της Τουρκοκρατίας, Athens 1986, 107-124.

Abstract
The aim of this paper is to address an issue regarding the digital resources created for the Humanities by both Signum-SNS and INSR. This issue focuses on the fact that these digital resources are not fully exploited by users. As these resources are created for Humanities scholars, CRIBeCu (now Signum) has started a pioneering work by means of a synergy between humanists and computer scientists. This has allowed cutting-edge research to proceed along with applied research and attention to users. Furthermore, humanists have suggested IT research and they have received, in turn, inputs from informatic results.
It can be claimed that the major potential of digital resources lies in their flexibility, although such a flexibility implies an high level of complexity. Despite the facilities put at disposal of users, the latter are discouraged by difficulties involved in the use of them.
Two typologies of resources have been created:
- digital collections for XML documents' search and consultation (e.g. BIVIO)
- collaborative tools for XML documents' management and advanced search (e.g. TauRo)
As statistical analysis demonstrates, users approach digital resources according to a traditional perspective: digital tools are consulted as a digital reproduction of paper documents, while the tools specifically developed for text management and analysis are disregarded. This approach enables users to exploit only the basic functions of the digital resources so that performance and fruition are less effective.
In this paper we will examine the reasons determining this phenomenon in order to develop a strategy which can contribute making digital resources more effective.
Keywords: digital library; search engine; XML document; collaborative tool; Renaissance

Introduction

Signum [1] (formerly CRIBeCu) is a computer science laboratory of Scuola Normale Superiore (Pisa) that provides and designs digital resources for specialists in the Humanities. It includes a team of humanists and computer scientists who cooperate with Istituto Nazionale di Studi sul Rinascimento (Florence).
The major purposes of Signum are digital humanities research, effective application of this research to digital resources and, finally, exchange of ideas with similar research groups.
The results of the activities carried out by this laboratory can be evaluated through an examination of the feedback from users.
Accordingly, this paper will present an inquiry made by Signum about the attitude of users towards two main typologies of resources, exemplified by the following projects: BIVIO – Virtual library online [2]; TauRo – search and advanced management system for XML documents [3], which have been developed by Signum.
The results of this inquiry point to a limited exploitation of digital resources. This indicates that either users are unable to work with these tools or that the latter are inadequate.
A detailed analysis of the actual needs and capabilities of consumers is required to provide more effective tools. In particular, it has been observed that in recent years users tend to prefer new digital tools to the resources provided by computational linguistics.

Case studies analysis

BIVIO was created as a response to a need to have Renaissance texts available and searchable online. BIVIO can be seen as a model case study. The presentation page of the project may be quoted: "The purpose is to guide philosophical, historical, artistic, philological research to create a virtual library, able to offer rare texts in their more significant editions and translations, made available thanks to adequate IT systems that guarantee multi-level information retrieval: from the

Chiara Cirinnà and Maurizio Lunghi (edited by), *Cultural Heritage on line: Empowering users: an active role for user communities. Florence, 15th - 16th December 2009* ISBN 978-88-6453-184-7 (print) ISBN 978-88-6453-187-8 (online)

easier, as words frequency, to the more sophisticated, apt to analyse the content". This statement clearly demonstrates that the purpose of the project is to stimulate text analysis through specific research tools. It is also stated that "the project offers aids (e.g. quotations lists and iconographical corpora) to a deep comprehension of the period in question".
On the website of BIVIO a resource access has been arranged which is parallel to those of libraries catalogues, and provides textual documents. An IT system has also been developed, which enables to analyse and compare texts in an innovative way. This system provides information which differs from any 'paper-like' approaches, such as visualisation of occurrences in the text retrieval results (snippet lists), which allows comparison between different occurrences of the same word and text search based on both two distant words and different variants of the same word. Nevertheless, it has been noted that these available tools are not fully exploited by users.
Let us focus on the reasons determining this phenomenon.
Analysis of access statistics indicates that BIVIO users are essentially humanistic operators. Indeed, the outer web-sites usually reaching BIVIO are mostly academic sites or sites which are concerned with themes treated in BIVIO (fig.1). Another evidence confirming a correspondence between the theoretical and the actual target of BIVIO regards the researches made by main search engines: 70% of them queries citations, works titles or specific authors (fig.2).

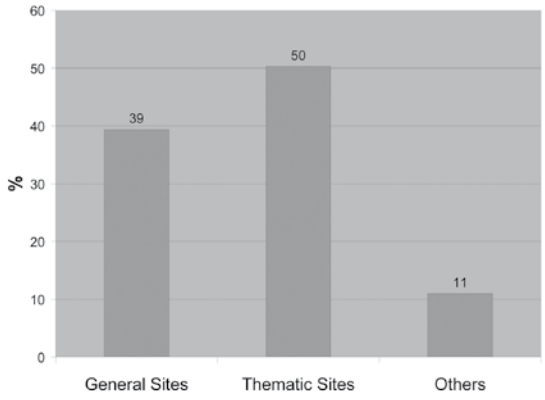


Figure 1 - Types of pages/URL linking to BIVIO

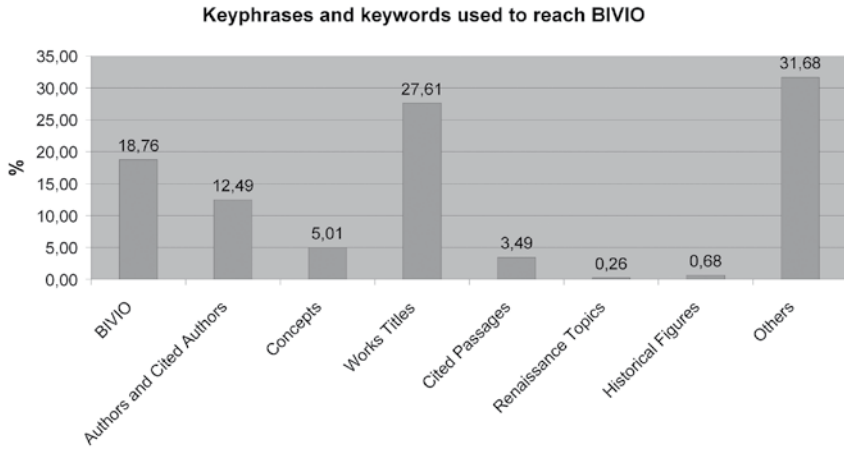


Figure 2 - Keyphrases and keywords used to reach BIVIO

Moreover, about 75% of users add BIVIO to their favourite links, which indicates an appreciation of this resource.
These data seem to demonstrate that BIVIO fully matches the needs which originally led to its creation, and that the users feedback is positive. However, a closer examination of these data rather indicates that users do not exploit the full potential of the tools provided by Signum.
Apart from the evaluation of any visits with a duration of less than 30 seconds (about 50% of the total), statistical analysis reveals that the interaction with the site reflects a traditional approach while the innovative retrieval tools appear to be scarcely used (fig.3).

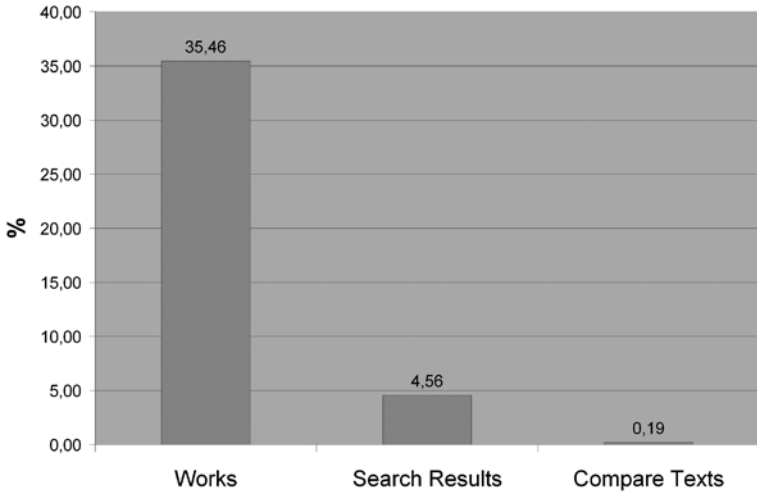


Figure 3 - Most visited pages

The results of this statistical analysis as well as the continuous interchange with the humanists may be helpful for interpreting the general picture above. Accordingly, some hypotheses can be advanced which have an important bearing on the promotion of original digital humanities research and new applications to cultural heritage:
users have a traditional working approach; generally speaking, they prefer to read texts while rarely asking for analysis tools.
users perceive information provided by BIVIO tools as something similar to paper data (rhyme concordance, word occurrence, lexicon, etc.) which are generally used for particular text analysis and for specific studies.
a specific training to use the IT tools applied to cultural heritage is lacking; as users may be not familiar with digital resources, they persist in using only the basic functions of these resources.
users do not trust research systems based on IT tools and so results are not considered reliable.
BIVIO does not fully satisfy Humanities scholars demands.
It has been suggested that hypothesis 5 can be easily verified through a direct inquiry by users. Statistics, however, lead us to discard this approach because users without any adequate training in managing these resources may not have a critical opinion.
Hypotheses 1 and 3 specifically refer to users and their training while hypotheses 2 and 4 are connected with the nature of the tools under examination and their appreciation by users.
In order to settle these problems, it is needed to promote the dissemination of digital resources such as BIVIO through a specific training for users. Furthermore, generalist systems may suggest to adopt strategies developing an easy access to digital resources, which maintain a high, scientific and reliable standard of the product.
Differently from a digital resource such as BIVIO, Signum has projected TauRo not in view of the actual needs of users, but rather by means of a new approach typical of Web 2.0. Therefore, a collaborative tool for XML documents' manage-

ment and advanced search has been created, which is able to exploit all the capabilities of a search engine previously realized by Signum: TauRo-core.
The idea of bringing together a community of XML documents users was substantially unsuccessful because a very few users have been involved. As a consequence, collaborative tools are not currently used and users limit their access to public resources consultation (according to TauRo statistical data).
Moreover, TauRo is not able to test the capabilities of the search engine because users do not use its advanced functions and only scroll documents or make simple queries.
Users show a similar attitude to two different digital resources such as BIVIO and TauRo for similar reasons.
Furthermore, the collaborative nature of a resource like TauRo leads us to focus on this specific aspect. Users load a very few documents, create a very few collections, and essentially do not share them. The following hypotheses can be put forward to explain this phenomenon:
XML format is not familiar amongst humanists.
XML format is used by scholars who do not like to share their own work.
TauRo system is too complex for users.
All these hypotheses suggest that a general solution for these problems is to simplify TauRo. This is possible by making TauRo even more complex and by developing a system that, thanks to a wide range of services, will make it more accessible for a larger audience and will attract specialists, who will be encouraged to use it as a working platform. Signum will also try to make TauRo interface more user-friendly. Indeed, in the future, users will be able to load documents in more common formats, which will be converted into XML format by the system, so as to preserve all the capabilities of the search engine. TauRo will also be provided with proper tools to analyse, note, mark, correct and edit loaded documents.

Conclusions

Analysis of users' attitude is important in order to develop new strategies for future applications, and to open a new path toward both humanistic and computer science research.
In this paper, we have focused on some problems regarding the use of two model digital resources realized by Signum. Furthermore, some solutions have been suggested on the basis of the key idea that digital humanities research does not simply target users demands, but it also helps to acquire new skills and to master new working methods.
Users training needs to be urgently increased for a more effective and aware approach to digital resources, while the latter should be more accessible and useful.

References

[1] http://www.signum.sns.it/
[2] http://bivio.signum.sns.it/
[3] http://tauro.signum.sns.it/

Abstract
BAM – the joint portal for libraries, archives, museums in Germany intends to become a single point of access for cultural content and serves users who do not want to search several different databases at different servers using different search interfaces and vocabularies for access. In addition to combining different information services from different institutions in one point of access, BAM can also serve as a portal for a single institution's libraries, archives, museums and media centres. BAM also tries to increase the visibility of the digital objects in the collections of the participants by cooperating with Wikipedia Germany and enriching articles with a link to content in BAM.
Keywords: Cultural heritage, portal, museums, libraries, archives, access

Introduction

When looking for digital cultural heritage information, users do not care whether the information they require is stored in a library, an archive or a museum [1, 2]. In the digital realm it is no longer relevant whether the original materials that are now available in a digital form were stored in a library or a museum or an archive [3]. The current development of libraries, archives or museums goes towards a digital memory institution where the information of all institutions is available online. BAM – the joint portal of Libraries (in German: Bibliotheken), Archives, Museums intends to set up such a digital memory institution for Germany providing a single point of access to users who do not want to search several different databases at different servers using different search interfaces and vocabularies. Such a single point of access is a major improvement because in Germany does exist a lot of digital resources but they are scattered all over the Internet like islands in the sea. In order to find these materials, the users have to know that these islands of digital materials exist, where they are located and what kind of resources they hold. So the users have to do some island hopping in order to find the information they are looking for. In addition, to access such a treasure island, they need to know the magic words Open Sesame as in Ali Baba's tale in One Thousand and One Nights, i.e. they must understand the various interfaces, know the right terminology and the underlying indexing structure for the database for each and every information resource. From the users' perspective it would be more effective and convenient to have one platform where they can stop and search all the available online databases - a single point of access.

BAM – A Joint Portal for Libraries, Archives, Museums

BAM (Fig. 1)[4] started as a project funded by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) in 2001. Since 2007 a consortium of library, archive and museum institutions hosts the BAM portal, among them the Bibliotheksservice-Zentrum Baden-Württemberg (BSZ), a library service centre that hosts the portal. At the moment BAM contains more than 40 million digital records contributed by several major German academic libraries, by sixteen museums and museum networks, and several major archives (cf. Table 1).
The BAM portal offers the participating institutions a joint cross-institutional platform for digital catalogues, repertories, and inventories. Therefore, metadata of the participating institutions are collected, stored, indexed and made searchable on the BAM server, while the media content, i.e. the digital materials such as images and – in theory also text, audio and video, is stored in the online databases of the participating institutions who keep full control over and responsibility for their digital materials using BAM only as a gateway and as a means to increase their visibility on the Web by contributing to large digital collection that attracts user traffic. For smaller institutions without an online database of their own, a hosting service is offered by BAM. Such smaller institutions can store both the metadata and the media content of their digital collections in the BAM database which allows them to present their content on the Internet without having to maintain a complex web presence including an online database. As a bonus for sharing their content via BAM, these institutions can include a search form on their websites in order to present their own content on their own homepage. This option is important for institutions with limited resources.

Chiara Cirinnà and Maurizio Lunghi (edited by), *Cultural Heritage on line: Empowering users: an active role for user communities. Florence, 15th - 16th December 2009* ISBN 978-88-6453-184-7 (print) ISBN 978-88-6453-187-8 (online)

BAM total number of digital records	41 195 322
Libraries	37 175 528
Northern German Union Catalogue GBV (some 330 scholarly libraries)	~20 M
Southwestern German Union Catalogue SWB (some 1200 scholarly libraries)	~13 M
State Library of the Prussian Cultural Heritage Foundation, Berlin	~3 M
Central Index of Digitized Imprints (ZVDD)	~0,5 M
Archives	2 905 652
State Archives of Baden-Württemberg	1,7 M
State Archives of Hesse	0,8 M
Federal Archive of Germany	88 K
Municipal Archives (Freiburg, Heilbronn, Reutlingen, Mainz)	86 K
Museums	291 563
Architecture Museum of the TU Berlin (collection of technical plans and drawings)	69 K
Historical Museum of the City of Leipzig	141 K
The Prussian Cultural Heritage Foundation, Berlin	11 K
digiCULT Schleswig-Holstein	18 K
Foundation Haus der Geschichte, Bonn / Leipzig	6,5 K
German Historical Museum, Berlin	6,5 K
Other sources (Kalliope portal)	822 708

Table 1: The total number of digital records in BAM

Figure 1: The BAM portal

To the present day, BAM is the only German cultural heritage portal on a national level as the German Digital Library (Deutsche Digitale Bibliothek, DDB) is still under construction and is not going online before the end of 2011. Therefore, BAM is currently a single point of access for all users who are searching items of cultural content on the German Web. As a consequence, the potential range of users is very broad, the major target audience being scholars, students, but also a general public of interested laypersons. As it is considered a central educational and scientific resource, access to the portal and the content of the participating institutions is free of charge.

BAM Local - Uniting Different Branches of an Institution in one Portal

Apart from serving as a portal for different institutions, BAM is also applicable for an individual institution or a city or region who wants to make accessible its digital collections from different branches such as libraries, archives, museums, photo libraries and media centres at a single point of access. The so called "BAM local" presents a single institution's or city's or region's collections from different sources in a single portal and in this way creates a single point of access for potential users.

The advantage of a "BAM local" application is obvious: most institutions or cities or regions maintain different information services which can only be accessed from individual Web-based applications such as Online Public Access Catalogues in one or many libraries, from search engine interfaces of different Web-based database applications in museums, archives and media centres. With "BAM local", all these different content providers can unite their collections in one metadata database with a single index and interface. The "Google slot" of BAM can be integrated into almost any Web design by a simple HTML form and the user will be transferred to the BAM results page which can also be adapted to the institution's or city's or region's corporate design. In this way, "BAM local" is applicable for many purposes.

Increasing Content Visibility by Collaborating with Wikipedia

In addition to serving as a central point of access, BAM tries to increase the visibility of the digital content of all participating institutions by collaborating with Wikipedia Germany. In August 2007 an alliance was formed that allows Wikipedia users to connect the encyclopaedia's web links section to a predefined query in BAM using a specific BAM Template (Fig. 2). Both information services can take advantage of this alliance: Wikipedia Germany offers its users a wide range of sources to investigate and BAM increases the visibility of its partners' digital content and draws traffic to their Web sites. Until December 2008 more than 900 BAM links have been created in Wikipedia and the process goes on, continually increasing the number of links.

Figure 2: The BAM template in Wikipedia

BAM and its Users

A detailed analysis of log files has not yet been carried due to lack of time and personnel. Hence the above mentioned target audience of the BAM portal has to be investigated further. The results of a preliminary examination of the BAM log files shows that there are more than 1 000 visits per day or around 30 000 visits per month (form June 2008 to May

2009). These numbers are small compared with those of major search engines, yet it is a reasonable start and a point from which to continue to build a stable and large BAM community. Especially the link to Wikipedia has increased the traffic considerably as the current examination indicates.

Conclusions

BAM – the joint portal for libraries, archives, museums in Germany intends to become a single point of access for cultural content on the German Web. In this way, BAM serves users who do not want to search several different databases at different servers using different search interfaces and vocabularies for access. To do so, BAM combines the different online information services from different institutions in one point of access. In addition, BAM can also serve as a portal for a single institution's libraries, archives, museums and media centres by combining their digital collections in one index under one search interface that can be integrated into the institutions corporate design. Apart from this, BAM also tries to increase the visibility of the individual digital objects in the collections of the participating institutions by cooperating with Wikipedia Germany. A Wikipedia template containing a predefined query to BAM can be added to any Wikipedia article and enrich it with a link to media content in BAM. Therefore, from our perspective, BAM is a successful tool to empower users who are looking for digital cultural heritage content on the German Web.

References

[1] Hedegaard, R. (2003) Benefits of Archives, Libraries and Museums Working Together. In Access Point Library: Media - Information – Culture. Proceedings of the World Library and Information Congress: 69th IFLA General Conference and Council in Berlin, Germany, August 1-9, 2003 <http://www.ifla.org/IV/ifla69/papers/051e-Hedegaard.pdf>, accessed: 09/29/09.

[2] Martin, R. S. (2003) Cooperation and Change. Archives, Libraries and Museums in the United States. In: Proceedings of the 69th IFLA General Conference and Council, August 1-9, 2003, Berlin. 1-10. <http://archive.ifla.org/IV/ifla69/papers/066e-Martin.pdf>, accessed: 09/29/09.

[3] Kraemer, H. (2001) Museumsinformatik und digitale Sammlung. [engl.: Museum Informatics and Digital Collections.] WUV-Universitäts-Verlag, Wien.

[4] <http://www.bam-portal.de>, accessed: 09/29/09.

[5] Kirchhoff, T.; Schweibenz, W.; Sieglerschmidt, J. (in print) Archives, Libraries, Museums and the Spell of Ubiquitous Knowledge. In: Archival Science – Special Issue on Digital Convergence in Libraries, Archives and Museums. Springer.

Abstract
The National Archives of Estonia has intensively digitized its most used archival units and developed internet solutions for presenting these materials free for everybody online since 2005. Among other solutions there are opportunities for users to add value to the digitized materials. Users can enter names of people described on pages of archival units to facilitate searchable access to mostly church books. Users can create custom online databases about the content of archival units and make accessible for everybody. Users can interact with each other in the forum and point to an exact spot on a page to ask help for reading hand-written text. Users can create their own link collections of digitized pages and spots on pages. Experience of the National Archives of Estonia says that users should be involved in time consuming processing of digitized materials for adding searchable data to them.
Keywords: digitization of archival materials, user involvement, content enrichment, online access

Results of digitization of archival records

The National Archives of Estonia (NAE) has digitized around 50 000 archival units with 5 million images comprising around 5 Terabytes of data during 2005 to 2009. Most of these archival units are church books from the 18th to 20th century that are the most used materials by genealogists in the Estonian archives. By the end of 2009 almost all Estonian church books are available online.

The essence of genealogical research in Estonia has changed radically since 2005 when the first digitized materials where published online at the environment of digitized content of NAE called Saaga (www.ra.ee/saaga/). Until then genealogy was a hobby for a few who could allow themselves spending time in the reading rooms of archival buildings mostly during working hours but also Saturday mornings and scrolling through the microfilms of church records.

When Saaga environment was opened genealogy became a popular way of spending free time for many people. All one needs for it is an average computer with an average broadband internet connection. Saaga is available for everybody 24 hours a day. Only free registration and afterwards logging in is needed for access. A drop-down of 25% of the quantity of physical users in the reading rooms was a logical result for these developments.

The other issue that has severely influenced the behavior of archival users and the character of usage of archival materials is the mass input of headings of archival units, series and archives into the archival information system (http://ais.ra.ee/). The input process was started in 1999 and the database was published online in 2004. By the end of 2009 all the headings (about 8 million) will be inputted and be available online – Estonia is then amongst the few countries where 100% of the archival descriptions are digital.

Since 2004 users make their searches in the archival information system instead of paper records. This has changed the variety of archival units used – new groups of archival materials are accessed that where practically not used before because of low knowledge amongst users. Users also spend no more time searching for the archival units in the reading rooms as they mainly do it online and order documents to the reading rooms before their visits.

Options for user involvement

As the amount of digitally available archival descriptions and digitized images has risen steadily and there are much more archival users than before, the amount and potential of the user community has grown massively. Genealogists have nonprofit amateur unions where they share knowledge and ideas and discuss different problems in forums (e.g. www.isik.ee/foorum/). The number of genealogists and other online users outnumbers the amount of archivists and IT developers in the archives a lot. Most of the users are waiting eagerly for every new piece of digitized documents and each new digital description. Several of them are true fans of archival and genealogical studies and have expressed their wish to help NAE in the process of making archival documents digital.

Chiara Cirinnà and Maurizio Lunghi (edited by), *Cultural Heritage on line: Empowering users: an active role for user communities. Florence, 15th - 16th December 2009* ISBN 978-88-6453-184-7 (print) ISBN 978-88-6453-187-8 (online)

As the archives can not let volunteers do the basic digitization work which involves physical scanning procedures, there are opportunities to involve users in the areas where the archives will probably never have enough resources to do the work. These areas have been described in NAE as the following:

- quality checking of digital archival descriptions for printing errors and logical faults,
- helping other inexperienced users in understanding the content of archival documents,
- describing the content of archival documents in a structured way,
- collecting similar data from different archival documents and making these thematic databases available for public.

REALIZATION OF USER INVOLVEMENT IN ONLINE TOOLS

In connection with user groups several tools have been made by the archives' IT developers to satisfy the users' wish to contribute to the digital content and the archives' need to have more searchable data about digitized documents of good quality.

Quality checking of digital archival descriptions

In the web portal of the Estonian archival information system (http://ais.ra.ee/) there are headings and other descriptions of about 8 million archival documents, series and archives.

As this data has been input manually during 10 years, several data has been transferred from legacy databases and the descriptions are in Estonian, German and Russian which use different alphabets then it is a known fact that there are quite many typing errors, data transfer errors, logical descriptive errors etc in the system.

NAE has launched a simple solution to allow users to give feedback about descriptive data that is not correct. The feedback button "Report a mistake" includes the number of the archival unit and other technical data with the user's message about the mistake and it is sent to the database administrator of the system for correction. Every day several mistakes are reported.

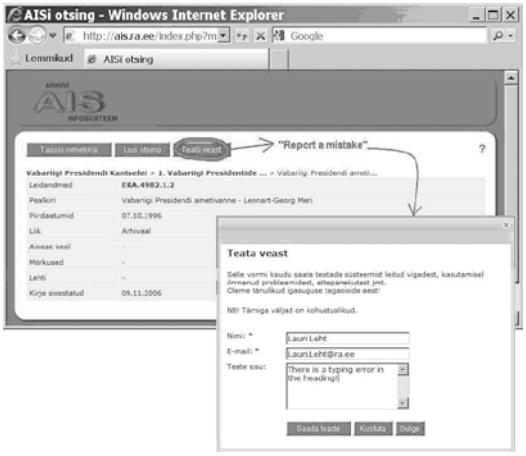


Figure 1. Reporting form for mistakes in the archival descriptions.

Helping other users understand content of documents

Most of the digitized documents are church books that are up to 300 years old, lots of them are hand-written in German Gothic writing which is quite hard to understand for inexperienced users.

Fortunately there are also several experienced users in the community who have been dealing with the church records for a long time and can give answers for most of the puzzles. As it was possible to free the archivists from the need of giving this kind of explanations online, NAE implemented in its Saaga environment (www.ra.ee/saaga/) a solution where users can select one or several areas on a digitized image and post them to a forum where volunteers from the genealogical society are eager to help each other in understanding the meaning of badly-written phrases.

Users can also combine the area selection function and the personal link collection function in Saaga and save their necessary data from the digitized images pointing exactly to the relevant parts on the image.

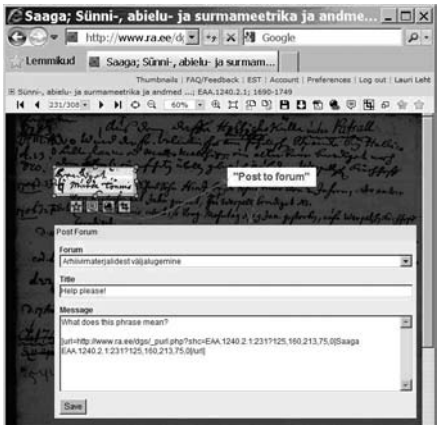


Figure 2. An example from Saaga of cropping and pasting areas of an image to forum.

Describing content of archival units in a structured way

The images of digitized archival units are raster images where no layers or text is optically recognized. This will probably be so for some more time as the OCR techniques for old hand-written texts are not yet practically available. Therefore NAE has created and given to the volunteer users a tool for indexing data of names of people from the church books as names are the most used search words and also the real essence of church records (www.ra.ee/dgs/

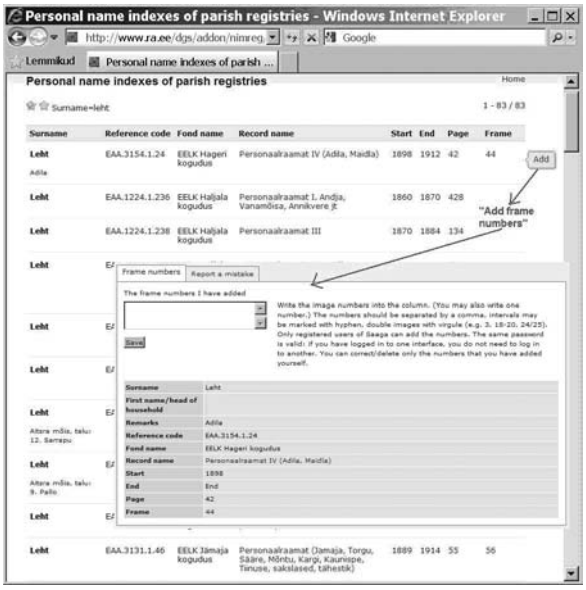


Figure 3. The online tool for indexing personal names of the parish registries.

addon/nimreg/). The genealogical society is doing the work of inputting names with page numbers from the digitized images. All the users of Saaga can help in connecting the church books' page numbers with the actual digital frame numbers as these figures always slightly differ.
As a result genealogists reach the pages of church books where the surnames that are of most interest to them are represented in a quicker way.

Creating thematic databases
Users of archival resources usually make some kind of personal databases about the topic that they are exploring, whether it is simply page numbers of important data or some highly structured specific data. These databases may involve similar data from different archival units or from several pages from one unit. Sometimes these databases are of more than just personal value for the user and the author may wish to expose these data for the general community of users so that others do not have to duplicate this work.
NAE has created in its virtual reading room solution (www.ra.ee/vau/) a possibility to create, manage and publish personal databases about all the possible archival content. Users can connect the rows in the database with digitized images in Saaga environment. Most of the databases that are published there deal with some kind of indexing of the pages of digitized archival units.

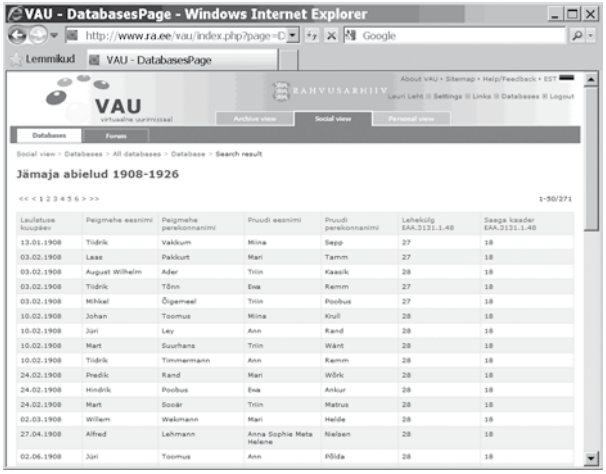


Figure 4. Example of a database made available by one user to the general public.

CONCLUSIONS
Experience of the National Archives of Estonia says that users should be involved in time consuming processing of digitized materials for adding searchable data to digitized images. If the public archives give convenient tools to users for that, volunteers from the user community are eager to start producing and publishing data that adds on to the digitized images. The role of the archives in the near future should be digitizing their documents according to popularity and listening to the users' needs for providing good tools for archival fans for creating added value.

Abstract. Today, archives and libraries are involved not only in digital projects, but also in definition of new policies in order to guarantee a long-term preservation of digital objects.
Digital preservation is considered a process that requires use of the best available technology and related procedures. In particular, information must be intact and readable from storage media; contents have to be accessible and interpretable; standard formats and migration plans must be developed. Digital data are stored in magnetic and optical supports which have different characteristics and life expectancy. National and international scientific committees promote standards and technical strategies to extend the useful life of digital media and protect them from degradation and technological obsolescence.
In this paper structure, technology, and degradation processes of common optical discs (CD, DVD, and Blu-Ray Disc) for digital preservation are described, with particular attention to Holographic Versatile Disc (HVD), an innovative technology which offers a storage density more capability than the other optical media. Furthermore, internal and external factors that can attempt the integrity of supports and data such as instability of components, environmental factors, and uncorrected handling are discussed. Finally, standard storage conditions and care for long-term preservation are reported.
Keywords: optical storage, holographic data system, recording materials, digital preservation device.

Introduction

In the last years organizations involved in preservation of digital information need to high reliability systems. Data can be stored on each medium that can represent their binary values (bitstream), such as magnetic or optical media. It is important to have knowledge of the different media, of particular software and hardware equipments for access and storage, and of conditions requirements for preservation.
Optical discs, due to their easy of use, large capacity and low costs, are considered supports for storing digital information. In 1982 the Compact Disc becomes the most common media for recording data. Ten years later, Digital Versatile Discs, increasing the capacity, had preferred. In 2007, Blue Ray Disc provided 25 GB.
Anyway, the lifetime of these supports, in other words the period of time in which the information is stored in safety, is matter of studies in all over the world. Data are vulnerable to loss and corruption; in fact, optical media are sensitive to heat, humidity, pollutants or can fail because of faulty reading/writing devices.
After some years, optical discs change their capacity, features, logical and physical format, and, consequently, hardware/software systems. In order to contrast digital obsolescence, contents must be copied periodically on new media and formats (refreshing and migration) [1].
Today, holographic supports are a new technology that promises to revolutionize the storage systems (500GB). In the past, the realization of holographic system has been discouraged by the lack of availability of suitable components, the complexity of holographic multiplexing strategies, and the absence of recording materials with satisfying optical storage requirements. Recently, new studies and researches have rekindled the interest to this technology.

Structure and technology

Optical discs, that use laser technology for storing and retrieval data, can be classified as follow: Compact Disc (CD), Digital Versatile Disc (DVD), Blu-Ray Disc (BR), and Holographic Versatile disc (HVD) [2][3][4].
CDs, DVDs and BRs consist of same basic materials and layers, but they are differently manufactured. There are many kinds of optical discs; the attention will be focused only on –ROM (read only memory) and –R (recordable) [5]. –ROM and –R discs have a multi-layer structure (Fig.1).
The substrate is a polycarbonate which provides the transparency useful for laser to reach the reflective and data layers. It also offers the necessary depth to maintain laser focus, and, at the same time, enough strength to remain flat.

Chiara Cirinnà and Maurizio Lunghi (edited by), *Cultural Heritage on line: Empowering users: an active role for user communities. Florence, 15th - 16th December 2009* ISBN 978-88-6453-184-7 (print) ISBN 978-88-6453-187-8 (online)

The data-layer contains digital information: in –ROM discs, data are "pressed" in the reflective/substrate layer (molding process); in –R discs, data are written by a high-power laser which changes chemical structure (burning process) of an organic dye (cyanine, phthalocyanine, azo based). DVDs and BRs can have one or two data-layer.
The reflective-layer is a metal which reflects the laser beam to the photosensor. Three types of reflective metals are normally used: aluminium, silver, and gold. The photosensor transforms optical into electronic signals and, by means to an analogue-to-digital converter (ADC), digital information is reconstructed.
A very thin lacquer is applied to protect the disc from exposure to the environment (protective-layer). An optional label is useful as top layer for graphics design and logos.

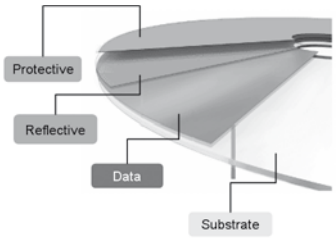


Fig.1 - Cross section of an optical disc

In all kinds of optical discs, data are marks (pits) impressed on the flat surface; the area between two pits is called land. An optical disc contains a track of pits arranged in a continuous spiral running from the inner circumference to outer (~5 Km). The drive reads marks on the track using a laser which measures the amount of light that gets bounced back from it. Areas with pits reflect the light less strongly than land areas. When photosensor detects a switch pit/land or land/pit, the system reconstructs the digital pulses. The pits on the data-layer are the physical manifestation of a complicated encoding process including multiplexer, interleaving, parity, error correction, modulation (EFM) [6][7].
CDs offer storage capacity 0.7 GB about, DVDs 4.7 GB for each data layer, and BRs, which use a blue-laser and achieve a spot size of a few hundred micrometers, provide 25 GB for each data layer (Fig.2).

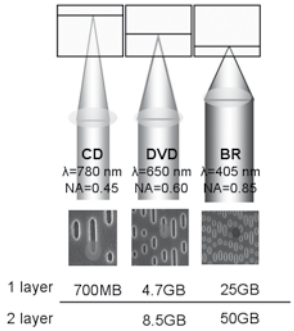


Fig.2 - Capacity, numerical aperture (NA), wavelength (λ) in optical systems

Holographic Versatile Disc (HVD), that use an innovative technology, is composed by a recording-layer between two substrates (polycarbonate) with in the middle a dichroic mirror that reflects the blue-green light and allows the red light to pass through in order to gather servo information. The servo monitors the position of the read head over the disc (Fig.3).
Recording-layer materials are divided in two classes: inorganic photorefractive crystals and photosensitive organic polymers [8][9][10]. Two optical techniques for recording data in holographic systems are used: two-axis (angle multiplexing) and collinear (shift multiplexing).

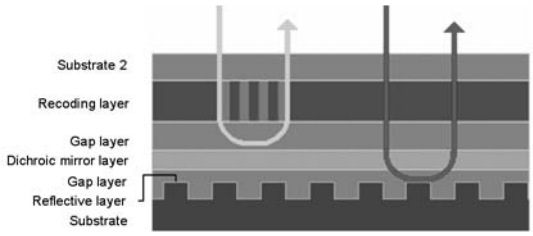


Fig.3 - Cross section of an HVD

During the HVD writing process, in the two-axis technique, binary data are disposed in a bi-dimensional organization (page). A spatial light modulator (SLM), or page composer, translates page into an optical pattern, called image, where ones and zeroes are represented as opaque (black) or translucent (white) areas; each area is also called pixel. Based on liquid crystals, the SLM offers a valid contrast and rapid switching between black/white states. At the moment, the page composer is structured as a 1024x1024 pixels matrix (pixel size ~15-20 micrometers).
Once the image is created, a single laser beam is split into two: information beam, which is directed toward the SLM, and reference beam, which is directed, using lens and light deflectors, into recording-layer. When the information beam passes through the page composer, portions of the light are blocked by the opaque areas of the image, and portions pass through the translucent areas. In this way, the information beam carries the image and when the reference beam rejoins on the same axis, a pattern of light interference, the hologram, is recorded in a light sensitive medium. By varying the reference beam angle, the wavelength, or the media position, many different holograms can be recorded in the same volume of material. This process of superimposed holograms, called multiplexing, yields the enormous storage capacity.
In the HVD reading process, the reference beam is incident on the medium under the same conditions used for recording and it produces a diffracted beam representing the image. The optical information is revealed by a detector array (CMOS or CCD) which allows extraction of the page from the measured intensity pattern. Then, the signal enters into the threshold, error correction and demodulation circuits; finally, the calculator can process the bitstream.
In collinear technique, reference and information beams come from the same SLM.
HVD capacity is 300 GB (Fig.4) about and the collinear strategy is used in order to guarantee storage information with simple and minimal devices.

Fig.4 – HVD with (right) and without cartridge (left)

Technical details of optical discs are reported in table 1.

Optical discs	CD	DVD	BR	HVD
Maximum Capacity (GB)	0.7	8.54	50	1000
Data rate (Mb/sec)	0.15	1.35	36	1000
Wavelength	780nm	650nm	405nm	500nm
Numeric Aperture	0.45	0.65	0.85	0.65

Table1 - Technical details optical discs

DEGRADATION

Each layer of optical discs can degrade. The polycarbonate is a very stable polymer and it degrade slowly respect to the other layers.
The data-layer, in –ROM discs, is coincident with the reflective-layer; in –R discs, high temperatures and high humidity accelerate the deterioration process of organic dye. Also prolonged exposure to natural or artificial light can increase the degradation of data-layer, altering the chemical and optical properties of dye. Phthalocyanine seems to be the most stable.
The degradation of the reflective-layer depends on material: aluminium is subject to oxidation in contact with oxygen, pollutants and high humidity more than the other metals; silver reacts with sulphur dioxide; gold is very stable. The effects of oxidation are loss of reflectivity and, then, loss of readability.
The protective-layer has an unknown permanence due to not declared chemical formulation.
One of the aspects of physical deterioration is the different dimensional changes of layers in consequence of thermo-hygrometric conditions fluctuations. The outer layers are more vulnerable than the inner because they are subject to mechanical damages.
Scratches can attempt to integrity of discs, obstructing the correct read/write operations and, consequently, corrupting the data. Figure 5 shows two images of a CD with opportunely caused scratches. The effects on substrate and reflective layers are analyzed by equipment for quality tests of optical discs.
As regards HVDs there is not exhaustive information about their components and systems failure.

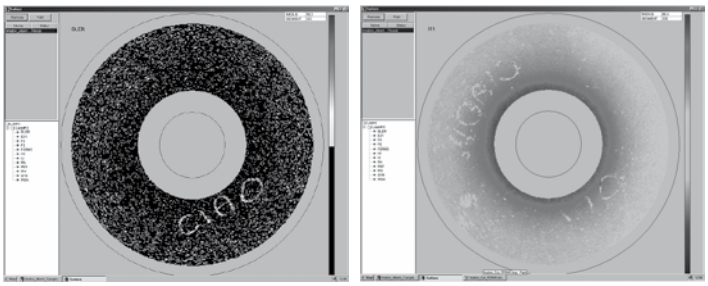


Fig.5 - Substrate (left) and reflective-layer (right) of a CD

STORAGE CONDITION, CARE AND HANDLING

ROM discs can be reliable for many decades if stored at suitable conditions, while CD–R, DVD±R and BR–R are at risk after just few years. Furthermore, degradation is expected over time, but some strategies can be taken in order to slow down it.
Generally, useful life of optical discs can be increased by storing at low temperature and low relative humidity, without fluctuations, minimizing the pollutants contents, avoiding the exposure to artificial and natural light, choosing proper shelves and boxes [11].
Storage temperature and relative humidity ranges recommended by ISO 18925 [12] are in Table 2.
HVDs present a protective cartridge in order to minimize effect of fingerprints and dust, but no information in order to guarantee a long-term preservation are reported in the scientific literature.

	Temperature	Relative Humidity
CD	<23	20%-50%
DVD	<23	20%-50%
HVD	?	?

Tab.2 - ISO 18925, Imaging materials - Optical disc media - Storage practices

Conclusion

Some standards describe methods for estimation of optical discs life-expectancy [13][14][15]; generally, it is possible to extend their life applying appropriate storage conditions, proper care and handling. In addition, frequent refreshing and migration are necessary in order to preserve the recorded information and to cope with the technological obsolescence. About the new optical supports, HDVs, it is possible affirm that they should become a candidate of next generation storage media but, at the moment, there is not exhaustive scientific literature about their systems, logical and physical structure, degradation, and their use in preservation field.

Reference

[1] K. Lee, O. Slattery, R. Lu, X. Tang, and V. McCrary, The State of the Art and Practice Digital Preservation, Journal of Research of the National Institute of Standards and Technology, Volume 107, Number 1, January–February 2002.
[2] ECMA-130, Data interchange on read-only 120 mm optical data disks (CD-ROM), 2nd Edition, Switzerland, June 1996.
[3] DVD Consortium, DVD-R for General, Part 1:Physical Specifications Ver .2.1, 1998.
[4] Blu-ray Disc Founders, White paper, Blu-ray Disc Format, General, August 2004.
[5] Optical Storage Technology Association, Understanding CD-R & CD-RW Technology, January, California, USA, 2003.
[6] ISO 9660, Information processing - Volume and file structure of CD-ROM for information interchange, 1988.
[7] G. Sharpless, An Introduction to DVD Formats, Deluxe Global Media Services Ltd, 2003.
[8] K. Buse, Holographic Recording Medium, Optical processing and computer, SPIE-International Society for Optical Engineering, 1998.
[9] A. B. Samui, Holographic Recording Medium, Recent Patents on Materials Science, Vol. 1, No. 1, 2008.
[10] Bayer AG, Long Life for data, Bayer research Magazine n°18, 2007.
[11] F. Liberati, G. Marinucci, M.T. Tanasi, Digital Preservation of Magnetic and Optical Support - Problems and Prospects, MATCONS, 2009.
[12] ISO 18925, Imaging materials - Optical disc media - Storage practices,2002.
[13] ECMA-379, Test Method for the Estimation of the Archival Lifetime of Optical Media, 2007.
[14] ISO 18921, Imaging Materials – Compact Disc (CD-ROM) – Method for Estimating the Life Expectancy Based on the Effect of Temperature and Relative Humidity, 2002.
[15] ISO 18927, Imaging Materials – Recordable Compact Disc System – Method for Estimating the Life Expectancy Based on the Effect of Temperature and Relative Humidity, 2002.

Abstract
This paper aims to bring a theoretical, methodological and phenomenological contribute concerning the interpretation by design culture on the ways of producing shared knowledge and culture and enhancing Cultural Heritage in the digital environment. In fact, design strategies, skills and techniques can be applied bringing a design driven innovation in Cultural Heritage digital exploitation. In particular, the experimentation of contents, languages and technologies, represents the most original approach in articulating a systemic model of shared knowledge accessibility and production available to the user.
According with this approach, the modalities of sharing and using culture and knowledge repertoires in the digital space respond to the complexity that characterizes both the cultural system and the communication one. Design culture proposes exploitation models and tools able to translate this complexity in contents and languages and to turn the available technology in virtuous devices enabling representation and access.
The systemic design approach basically introduces the direction of producing, using, managing, experiencing cultural repertoires on line, working both on the archetype of the catalog and new collaborative formats, giving shape to a wider concept of accessibility: from making "available" the Cultural Heritage to providing the opportunity for diverse community of users to use it in practice.
New formats are therefore designed, formats capable of responding to the emerging communication practices: solutions such as visual timeline are integrated with multimedia, video and with collaborative and customised tools, making them accessible and usable by the various communities of users (stakeholders such as professionals and researchers, but also the large public).
We analyzed these critical nodes through a phenomenological mapping of virtuous experiences and examples, able to identify the potentialities of Web 2.0 as a platform for an integrated communication system, which is able to re-orienting Cultural Heritage valorization towards social practices and converastions. Designing new paradigms of shared knowledge and culture production and use, we (as researchers and institutions) can move from the simple concept of accessibility to that of "use value".
Keywords: Design driven innovation, Use value, Storytelling, Performance, Place

"New" Cultural Heritage on line and dynamics of "use value" design driven

What is Cultural Heritage made of? It's commonly agreed, that, even if Cultural Heritage appears to be fixed and immutable, the concept has evolved by time. How human cultural artifacts become Cultural Heritage is a dynamic process, because value is not a technical quality embedded in forms and processes, but in the way they are integrated in the social lifestyles and patterns. The Cultural Heritage is the result of social relations, and increases its sense the more it is recognized and incorporated in the collective conscience of a community, in other words, "practiced" (1) in its "use value" (2).
Cultural processes require complex times of negotiation and settlement longer than the ones experienceable by a community, and it's necessary to split them in phases in order to make them synchronic and acceptable by people. For example, according to Dorfles (3), «art is a changeable reality whose meaning differs depending from time, and can be identified with myth, religion, society, technology»: this leads to different interpretations and fruition modalities, corresponding to the user, the context and the time. The processes of "genetic coding" of Cultural Heritage is not neutral: something appears unquestionably worth of cultural value, only under the beliefs and the socio-cultural constructions of an age. Consequently it's necessary a precise "investment" (for instance an enhancement project) to deliberately underline a particular content as valuable to the community: it is a specific will of social construction of a community Cultural Heritage. This is undoubtedly a selective and elective process of social production and reproduction of values and meanings, that ends with a distinctive collective attribution (1).

Chiara Cirinnà and Maurizio Lunghi (edited by), *Cultural Heritage on line: Empowering users: an active role for user communities. Florence, 15th - 16th December 2009* ISBN 978-88-6453-184-7 (print) ISBN 978-88-6453-187-8 (online)

In the contemporary society, the digital environment appears to be the most receptive context in enabling and incorporating the expression and legitimating of new heritage forms. In fact, it includes a wide repertory in consistence and typology of "new" Cultural Heritage forms and processes: from archives of digitalized tangible artifacts, to digital libraries of cultural expressions, catalogues of new forms of cultural production, and repositories of local knowledge, they all document the co-existence of the different formats that cultural identities can assume in the web. Databases and digital repositories has been explored in the last 20 years as the most recognised model of knowledge and cultural contents archive, but in the network age, the active role of the user, together with the obsolescence of data and the interoperability of formats, require to rethink this conceptual model in a more participative "locus" for the building and exchange of collective and visual identity of a territory and its community. Looking at the more recent examples, it is always more and more evident that the digital environment has changed into a "place" that facilitates the social and collective construction processes of the value of new heritage forms, and not only a "space" to store them. In this perspective, participative processes become cultural expressions too, and sometimes Cultural Heritage themselves.
In our opinion, this transformation has not been spontaneous: it has been consciously led and shaped by communication design, according to the emerging of a communication paradigm shift from availability to accessibility, from usability to participation, and as a response to the complexity of the contemporary cultural production system too. So, in the last 10 years the design of digital formats for Cultural Heritage enhancement has been addressed to experiment languages, technologies, collaborative and sharing tools, to enable those cultural negotiation and legitimating processes, apart from building a collective and shared memory: in other words, to play and act the heritage beside than document it. In this sense, communication design supports the dynamics of transformation of the Cultural Heritage value from "value per se" to "use value".
In particular, in the digital environment, the "use value" of Cultural Heritage relies on the capacity of design to enhance and make accessible the Cultural Heritage as a system and as a process for new different uses and users. The digital models and tools more design driven are the ones that apply the strategic and communicative potentiality of design to enhance and visualize the Cultural Heritage in a "re-usable" way, connecting its physical aspects with the digital ones. The first use value enabled by the design approach is generated by the exploitation on line of the heritage systemic nature, underlining the context and place where it has been generated from (from the physical localization to the natural, territorial, environmental, cultural and immaterial conditions which determined the "form" of the heritage and oriented its development), context that impacts and suggests new opportunity of fruition and further dissemination. The second use value enabled by design is arisen by the explicitation on line of the heritage process nature, enriching in its tangible elements with intangibile aspects, like abilities, skills, narrations, performances and procedures, useful for its innovative production and re-production.
The Cultural Heritage contents, in this systemic approach, are managed and processed by communication design as "products" enjoyable by the final user (experts or generic) and usable for the production of new cultural contents, or educational purposes.
From the following case study analysis, it will be evident that the new digital formats are designed as devices that empower the user by suggesting and enabling opportunity of practice, re-use and re-contextualization of Cultural Heritage, structuring in the web participative repertoires and tools for the bottom–up production and experience of knowledge and culture.

Communication design for digital Cultural Heritage contents direction

Turning every form of Cultural Heritage into an open resource by setting up digital integrated management systems is no longer the sole aim of communication design.
Its contribution, in terms of analytical and design tools needed to define one or more design models enabling the transformation of any cultural production into an open resource, is no longer based on cataloguing, on thesauri, on indexing, but focuses on the dialogue between digital atlases – complex and fluid communication systems based on information display – and the development of dense, localised systems, i.e. analogical atlases, based on narration, performance and places. If the places traditionally connected with knowledge in the field of the Cultural Heritage – let us call them nodes – are analogical, one may not think of managing large amounts of data and materials, of having all of the Cultural Heritage in one network. Hence the process started by Communication Design should go the other way round, i.e. it should not be based on the quantity, but on the quality of the cultural mediation that is made possible by

new technologies (databases, thesauri) and, even more importantly, new languages (information design, video, etc.). The relationship between the analogical and the digital, characterised by a mutual exchange, makes the communities of Cultural Heritage users dynamic, by connecting them to each other through Narrations, Performances and Places – seen as the new 'hotspots' of the analogical/digital dialogue.
Today the enhancement of this heritage, both material and immaterial, in the digital and analogical environments is not carried out only by means of digital archives, online collections or online museums, but also through processes like narrations, performances and places, which enable the individual and social production and re-production of Cultural Heritage knowledge.
The Cultural Heritage system can use Design to try and put in place this process, starting from the metaphor of Aby Warburg's meta-discursive and discursive atlas (early 20th century) in which he takes the ancient knowledge and images collected in his Library of the history of culture and displays them in a place like the Mnemosyne atlas, up to the Actor-Network Theory (4) which takes shape in the cartography of controversy. We may consider the two methods as fundamental research practices for the development of new Cultural Heritage systems, both of which are based on the generation of atlases into which converge various representations by the overlapping of several different levels. Be it an analogical (Warburg) or a digital (Latour's controversy sites) atlas, it is necessary to construct the toponymy of the maps of contents of the atlas – in our case concerning the Cultural Heritage – as an ethnologist would do, describing every object and every social fact as a Network.
The transition from the Mnemosyne Atlas to the Actor-Network Theory makes it possible to explore the ways in which past, present and future communities develop and maintain the connections between individuals and groups by means of narrations/mediations, performances/processes and places/systems managed by old and new languages (theatre, art, cinema, video, web). Studying all the actors (human being, technological artefact, institutional body, legal norm, etc.) who collaborate, more or less directly, to the creation of a (material or immaterial) piece of the Cultural Heritage is not enough, because everything depends on the type of action linking the various actors. On the one hand is the "existing" or "ready-to-use" Cultural Heritage; on the other the Cultural Heritage "under construction", from a state of "fact" or "artefact" to the "worknet", depending on the action of a vast network of actors. The work, movement, flow (action) that is generated is always an actor and, because actors and networks are two faces of the same reality, the search of new languages for the Cultural Heritage can no longer go into the direction of the digital alone, but it must also, above all, re-orient itself towards the analogical.
If we focus on the tree design-oriented actions/nodes of this digital-analogical dialogue a few questions are raised:

- Analogical/digital narrations. How can a community of users for a certain category of Cultural Heritage tell stories and define its own identity and link it to other identities? Investigation methods include storytelling, memory and the life story of the objects. These are cases in which strategies are applied to tell these stories outside archives and museums using digital technologies and new languages, e.g. the oral history of DHS (Design History Society) recording reminiscences, memories and experiences of the art community within the community of design history. The role of narration is meant as a strategy for emotional sustainability in a contemporary community of users and can potentially give birth to collaboration projects.
- Analogical/digital performances. How do the identities of the communities of users for the Cultural Heritage represent and manifest themselves? The performances of individual and collective multiple identities are considered, developing the Actor-Network Theory and the performativity of the Cultural Heritage, to build new spaces of expression, be them analogical or digital.
- Analogical/digital places. How do the communities of users act within and without space boundaries to preserve and create places for interaction and sharing? What roles do the communities play in the creation of places? The study of communities connected by way of digital networks, spaces and places will make it possible to identify the possibilities of innovation, developing competences and bringing about a greater social inclusion. For example, will it be sufficient to rest on cultural districts, producers of social and cultural inclusion in a territory and foundation for the building of a collective (analogical and/or digital existence of the Cultural Heritage?

Being interconnected, all the three actions bring about the construction of atlases which need new tools and new representation formats, new devices. The change takes place in language: for example, thesauri as languages need

to present contents in the form of narrations and performances through languages which are alternative to digital interfaces, e.g. oral story-telling, theatre and artistic performances. Conversely, places (cultural districts) need to present their contents – material, immaterial and environmental - by integrating them into cultural services targeted to users and to the development of relevant production chains using the new digital languages (video and collaborative tools).

Case study framework and interpretation

Below are therefore proposed case studies of each of the three actions (storytelling, performance and place) capable of triggering a virtuous relationship between analog and digital nature of Cultural Heritage, respectively: Oral History project by the Design History Society (www.designhistorysociety.org/projects/oral_history/index.html; www.vivavoices.org) and Telling Lives by BBC (www.bbc.co.uk/tellinglives) (5); on one hand the Digital Library by Sardinia District (www.sardegnadigitallibrary.it), because of the ability to collect different forms of expression and modes of representation of the identity of the Sardinian culture; on the other hand Cultura Italia (www.culturaitalia.it) and the database of the Powerhouse Museum in Sydney, Australia (www.powerhousemuseum.com), for the performative research by the user and his interaction with online resources; Monti TV, a Web TV in Roma (www.montitv.it) and the project Changing

		Case studies							
		DHS	Telling Lives, BBC	Sardegna Digital Library	Cultura Italia	Power-house Museum	AEC	MontiTV	Story-mapping
Communication aims	*Documentation*	√		√	√	√	√	√	√
	Education		√		√	√			
	Participation		√			√	√		√
	Collaboration	√							
	Promotion			√	√	√		√	
Users	*Experts*	√		√	√				
	Common people		√	√	√	√	√	√	√
Use value as:	*Storytelling*	√	√						√
	Performance			√	√	√			
	Place						√	√	√
Contents	*Tangible artifacts*			√	√	√			
	Cultural production	√	√	√	√			√	
	Local knowledge		√	√		√	√	√	√
Languages	Visual timeline		√						
	Visual mapping				√		√		√
	Video and multimedia	√	√	√	√	√	√	√	√
	Collaborative tools		√				√		√
Technologies	Streaming on line		√					√	√
	Library (On demand)	√		√	√	√		√	√
	Integrated devices			√ (podcast)		√ (podcast)	√		√ (mobile)

Linz and Wikimap Linz by the AEC - Ars Electronica Centre (www.aec.at; http://wikimap.hotspotlinz.at/de/index.php) (6); in the end, transverse to the narratives and places, we report the project Storymapping by the Center for Digital Storytelling (www.storymapping.org; www.storycenter.org). These cases were selected because they are representative of what is currently available online for access to Cultural Heritage and because of their potentiality for the use value, as it was described in previous sections, and for the communication language adopted. In particular, they allow you to link the Cultural Heritage with the dimension of time and history (timeline) or with the space and the reference area (mapping), while relations between actors in the community and the culture are developed through audio-visual narratives and the collaborative tools of Web 2.0.

Unfortunately, we don't have enough space to devote to each case a detailed analysis, but we can summarize in the table below some of the key features emerged.

We can focus on just certain elements with the aim of deriving the key factors in the proposal of design-oriented model for the cultural production and sharing.

The relationship between the scale of good (contents) and treatment (languages) is particularly evident in cases where the audiovisual and multimedia are used to enhance the narrative and emotional dimensions, re-producing knowledge and practices through audio-visual shot and, at the same time, producing new goods derived from the original ones (the document itself) and able to maintain and pass on new media that oral dimension typical of the cultural and tele-visual tradition. Orality is therefore not only a research methodology (cf. DHS), but also a documentary connotation of educational and popular communication. Moreover, Cultural Heritage visualization in relation to historical and geographical context, can convey information and widespread cultural knowledge from each single artifact to its relationship with the actors and the territory. The metaphors of the timeline and the map provide the user with a space of memory (7) triggering the performative dimension of interaction and the articulation of pathways for personal research and enjoyment. Digital technologies and the Web platform provide tools for georeferencing, while they encourage community participation and the production of content, on the other hand they tend to conform visual solutions and forms of representation.

In this regard it is useful to refer to the idea of Web as platform (8). Besides the technological feature of the distribution platform, it is necessary to consider the complexity of the environment in which it operates and the features of the Internet medium: the user is able to manage the information through a set of services, architectures of participation and collective intelligence. The scalability of digital content will allow the dissemination and use beyond the individual device access, as the Web was a single big software upstream of all devices and common to all nodes in the network. Starting from this premise, then, you can think of a new metaphor for interfaces and models of knowledge management: it's no more a personal desktop but a place to exchange stories.

Finally, it is important to distinguish between a communication addressed to a community of interests and practices composed of experts and a very large or local community, that can access to Cultural Heritage without expertise and may or may not refer to a specific geographic area. Indeed, the goals of communication are in close relation with the characteristics of users, as well as the tone and style of project is therefore consistent with both aspects. Communication design will identify, therefore, adequate communication solutions both in terms of languages and technology of fruition.

Conclusions

The systemic approach of design proposes, therefore, the cultural district as a system of goods and actors and as a communication system: within media convergence is useful moving towards integrated communications strategies that leverage the Net in order to articulate the Cultural Heritage communication into different formats and devices, that can produce value by multiplying the ways and contexts of use (from access to use). While it is true that the present condition is that the "always on", the Web is everywhere and represents an incredible potential for interaction, participation and collaboration (this order is a progressive path, from a simple access to an increasing specialized use). The cultural production traditionally understood as "high", faces diffuse and bottom-up practices, resulting in fertilization process of the cultural system and dialogue between actors in the district. In this sense, therefore, there cannot be cultural district which does not correspond to an integrated communication system. The design culture is proposing a participatory parardigma which is founded on one hand on mapping of cultural assets for their contextualization in historical, geographic and symbolic sense; on the other on the use value, understood as a re-appropriation and re-production of

the goods themselves through the acquisition of interpretative tools and dialogue. Communication design is thus able to integrate the strategic dimension to the forms of expression more suited to the process of translation of knowledge, consistent with the specificity of the actors involved and oriented to the strengthening of knowledge networks.

References

[1] Toscano M. A., Per la socializzazione dei beni culturali, in Sul Sud. Materiali per lo studio della cultura e dei beni culturali, Jaca Book, Milano, 2004
[2] Montella M., Valore e valorizzazione del patrimonio culturale e storico, Electa, Milano, 2009
[3] Dorfles G. , Le oscillazioni del gusto. L'arte oggi tra tecnocrazia e consumismo, Skira, Milano, 2004
[4] Latour B., Reassembling the Social: An Introduction to Actor-Network-Theory, Oxford University Press, Oxford, 2005
[5] Piredda F., Design della comunicazione audiovisiva. Un approccio strategico per la "televisione debole", FrancoAngeli, Milano, 2008
[6] Kuka D., ' Linz Changes. Under the Urban Microscope', in Stocker G., Schopf C. (edited by), Ars Electronica 2008. A New Cultural Economy. The Limits of Intellectual Property, Hatje Cantz Verlag, Ostfildern, 2008, pp. 103-107
[7] Yates F., L'arte della memoria, Einaudi, Torino, 1972
[8] O' Reilly T., What Is Web 2.0. Design Patterns and Business Models for the Next Generation of Software, 09/30/2005, http://oreilly.com/Web2/archive/what-is-Web-20.html

Abstract
At the intersection between the continuously growing demand for digital information and the necessary preservation of cultural heritage, digitisation is desirable – and maybe soon unavoidable – for many libraries, archives and other institutions in the field of information science.
Especially smaller and medium-sized academic specialised libraries face the challenge of digitally preserving their unique, heterogeneous collections of different materials and formats and at the same time satisfying the demanding academic needs of their users.
A solution to the problem is sought by a team of interdisciplinary expertise. Working on a conceptual study, the library at the Ibero-American Institute (IAI) Berlin collaborates with the technology experts of Fraunhofer-Institute for Production Systems and Design Technology (Fraunhofer IPK) and arvato direct services Wilhelmshaven GmbH.
The project is characterised by a comprehensive collection and analysis of all factors that are crucial in connecting digital preservation and user demands in the best way possible. These findings shall help conceptualise beneficial, innovative and flexible technical solutions and workflows for automated digitisation of two-dimensional printed cultural heritage, and will lead us away from the dilemma of being caught between user demands and preservation.
Keywords: digitisation, library, technology, preservation, user demands

A Conceptual Study for Comprehensive Digitisation Enterprises

The Ibero-American Institute is an interdisciplinary centre that combines research, culture and information: the institute employs and supports international scholars and regularly hosts cultural exhibitions and events; its library is Europe's largest specialised library on Latin America, Portugal, Spain and the Caribbean. The users of the IAI are national and international scholars as well as students. In its capacity of a special collections library, the IAI acquires material on special subjects as thoroughly and exhaustively as possible. As a consequence, the collections are in large parts unique as well as characterised by various materials, conditions, shapes and formats. What is more, special collections libraries have an archival function and mostly only buy one copy. Consequently, in theory, every item sooner or later is due for preservation, and thus desired to be digitised.
Due to its users, acquisition strategy and embedment in science and research, the IAI combines features of academic libraries, archives and other knowledge institutions and can therefore function as an exemplary institution in a study that focuses on finding beneficial, innovative and flexible technical solutions and workflows for automated digitisation of two-dimensional printed cultural heritage.
In order to generate such a concept and to profit most from the digitisation of the IAI's collections, the interests of both the material as well as the users of the library have to be acknowledged, evaluated and respected as much as possible. For this endeavour, a team of traditionally separated sectors combines their competence. Personnel of the IAI contribute their knowledge about library and information science, collections, user needs and every-day workflows. Fraunhofer IPK, among the world's leading experts in the field of automated virtual reconstruction of destroyed documents, and scanning services provider arvato services contribute to the evaluation of the technical status quo with expertise in the fields of digitisation, handling of original material and technical requirements and engineering. This expertise is complemented by experiences learned about in good practice reports, interviews with representatives of German digitisation centres that collaborate with libraries, as well as a survey conducted among selected specialised libraries, archives and other, similar institutions. Furthermore, certain standards that have already arisen in the still young field of digitisation, as well as legal restrictions for Germany, will be collected and considered for the representation of a generic digitisation workflow.

Collection Overview

Among the first facts to be collected – with the help of a questionnaire conducted in the summer of 2009 – were those about what materials are held by specialised libraries, academic libraries, archives and other knowledge institutions.

Chiara Cirinnà and Maurizio Lunghi (edited by), *Cultural Heritage on line: Empowering users: an active role for user communities. Florence, 15th - 16th December 2009* ISBN 978-88-6453-184-7 (print) ISBN 978-88-6453-187-8 (online)

Since our study focuses on two-dimensional material only, this list lacks items such as audio-visual media that of course also form important parts of these institutions' collections.
Still, it turned out to be quite a panorama of different media: monographs, bound and unbound journals, newspapers, loose-leaf-collections, posters, sheet music, folded and unfolded maps, microfilms, microfiches, photographs (paper, slides, glass plate negatives), postcards, single documents, press-clippings, files, time-tables, brochures, blueprints, certificates, official gazettes, periodicals (proceedings etc.), patents, correspondence, diaries, medieval manuscripts, papyri, portraits, autographs, inherited special collections, art prints, sketches etc.
As far as the IAI's collections are concerned, they consist of about 1,000,000 monographs, newspaper and journal issues, about 300 bequests as well as other special materials collections, which include

- about 900.000 press-clippings
- about 200.000 microforms
- about 72.000 maps (topographical maps, city maps, roadmaps, historical maps and thematic maps, e.g. geology, highway systems, land use, settlement studies, languages, borders, botany) of which about 6.600 were produced between 1851 and 1945
- about 70.000 manuscripts (correspondence, notes)
- about 60.000 photographs (plus about 22.000 slides)
- about 10.000 glass plate negatives
- about 3.800 posters
- about 2.200 postcards.

For the purposes of the study, the IAI's collections are to be characterised as thoroughly as possible. Due to the high number of bound material (monographs, journals and newspapers) and owing the fact that they form, after all, the typical group of library material, the collections at IAI were divided into bound media and special media, i.e. bequests, maps, posters, press-clippings and images. As far as the first, 'traditional' group is concerned, it was decided to take a sample of 500 items and to note their characteristics in detail. The special materials were examined in single groups and their most important attributes noted (material, number of items, size and format, storage, damage, indexing).

Decisions for digitisation – influencing factors

As has been established, any digitisation endeavour is, or rather should be, characterised by considering the various factors of preservation, user demands as well as the environment of the library.

Material

There are some media within the IAI's collections that qualify for preservation more than others, for example monographs and journals with acidic paper (about 15%). Also newspapers suffer from this lack in paper quality. Here even more, the paper literally dissolves when touched – some papers cannot even be taken out of the bundle. Furthermore, many items in the bequests (letters, scrap paper, photographs, and notes) are extremely fragile because of low-quality materials used and years of improper storage. The same can be said about the glass plate negatives: the layer of gelatine protecting the carbon motif is starting to come off and, moreover, they have been and still are unsuitably stored, so that their own weight is likely to crash them. Out of the photographs, many are bleached out or darkened, and on some of them a chemical reaction slowly renders the motif beyond recognition. As far as preservation is concerned, these would be the materials preferred for digitisation at the IAI. Apart from the mere scanning process, a digital removal of damages is of great interest.
Not necessarily damaged, but exceptionally rare are the IAI's bequests. Deceased scholars, some of them among the first Europeans to do research on Latin-America and its indigene cultures, have let their notes, scripts, drawings, photographs etc. to the IAI. On the one hand, this material draws scholars from all over the world to the institute, since most of the material is new to the scientific world. On the other hand, the collections are often in a bad condition, completely mixed up and not indexed at all – in other words, they do not qualify for material to be used in a library. Digitisation of these special collections would consequently mean both – preservation as well as preparation for scientific research. The presence of a digital copy would protect these materials from further handling; the originals could be stored away properly and only be offered for special research purposes.

While the general overview and the questionnaire showed that a wide range of materials exists, the sample of 500 revealed that even within a relatively homogeneous group (i.e. bound media), there is a wide range of different attributes that need to be considered – especially when having an automated process in mind: different paper qualities within one item (thickness, acid, size...), damages (yellowing, mould stains, tears, bends, acid, dirt, water stains...), aperture angle, attachments that are folded and bound within, print shining through, handwriting, gothic print or abnormalities in the layout (tables, landscape format, different fonts, irregular foliation).

Users

At the IAI, users with a scientific interest form the major group. These users, most of them employed researches or external scholars, work with all the IAI's materials, especially however with the bequests, of which they often are among the first to ever do research. On the other hand, many students of Latin-America related subjects use the library, since its collections are far wider ranged than those of the university libraries. These users mostly work with journals and monographs. For the project, IAI scholars and librarians established the requirements of both user groups regarding digitisation. To the scientific work of the scholars, it is important that...

- ... the digital copy is as authentic as possible (colour, fonts, notes should all be as in the original)
- ... all digital copies (also those of manuscripts) support full-text search
- ... (especially visual) media, e.g. maps, support additional functions such as digital navigation and connection with other material
- ... fragile materials are preserved
- ... rare materials and unpublished collections (if already properly indexed) are preferred in order to have them more easily and widely accessible
- ... the digitised material is presented in online platforms, since it would make the collections known, connect scientists and thus be fruitful to research.

Other users, mostly students and the interested public, consider being of importance:

- the content (as opposed to the shape, condition, feel...)
- a full-text search
- authenticity and integrity of the digital images
- a fast and easy access, since they often have a very limited time frame, preparing for an oral presentation, a report, a paper, a thesis
- excerpts: they often merely want a chapter, an article, sometimes only a few sentences and are happy to leave the heavy book in the library and only extract what they need, e.g. on portable storage devices or via e-mail
- a simultaneous presence of one work for when there are several users interested in the same subject (e.g. exam preparations)

Library

Apart from considering the demands of the users and the necessities regarding the materials, it is just as important to ask the librarians their opinion about the essential needs to satisfy both sides' demands as well as possible.
From the librarians' point of view...

- ...the digitisation process should be adaptable and integrable to the current book processing at the library, because only perfectly integrated does the process save time and further allow the library to keep relying on their own expertise regarding formal and subject indexing
- ... the process should interfere with the daily work (especially circulation and return) as little as possible
- ... there must not be any violations of copyrights
- ... automated formal and subject indexing as well as quality control during the scanning process would be ideal.

Standards and Legal Questions

It is advisable the digitisation process follow certain standards. In Germany, the Praxisregeln Digitalisierung by the DFG (German Research Foundation) serve as a benchmark in this case. They regulate criteria such as factors influencing what

to be scanned, file formats, generating full texts, organisation of the resulting metadata, authenticity, image quality (colour, size), storage (short and long-term), data exchange, integrity or presentation. Even though these standards are only binding for projects funded by the DFG, these rules cover many points and are constantly improved by practical users. Our study will consequently attempt to respect these rules in the conceptualisation of technical solutions. Digitisation can, of course, only be realised with material that does not fall under copyright. In general, the act of digitising is, just as the paper copy, an act of duplication and therefore only allowed under certain circumstances. For example, the point in time when a publication does no longer fall under copyright differs from country to country (in Germany, 70 years after the author's death). In general, attention has to be paid regarding who gets which materials where, in what number and for what purpose.

Conclusion and Next Steps

Digitising library material represents an intersection between user demands and preservation. This is especially true for specialised academic libraries and similar institutions, holding collections of various materials and formats. Digitisation improves the institutions' services in terms of a faster, easier access to documents that are fully searchable, and at the same time helps preserve rare and damaged originals. Different materials and their conditions as well as requirements of different user groups and the library will have to be considered and brought together, integrated into existing workflows, acknowledging established standards and legislation. All these factors have to be considered in the conceptualisation of new technology, and the above mentioned demands from the various perspectives set quite high expectations for the digitisation system, which should allow:

- flexible automatic digitisation of all printed materials in different formats
- interactive quality control
- identification and indication of expired copyrights
- an excellent image quality, suitable for OCR (Optical Character recognition) including manuscripts and gothic print
- easy generation of structured metadata through layout analysis modules (automated indexing)
- at least one master copy, and some commercial surrogates for publication, one representing the original and one full-text searchable version
- integration of the process into the existing workflow at the institution, including personnel, finances, logistics, presentation
- careful handling of sensitive material
- long-term preservation of the digital files
- and of course a reasonable price to be viable for the target institutions.

In the second phase oft the project, which ends in July 2010, a technology monitoring will be performed in order to identify new solutions and its capabilities, as well as possible opportunities for further development and research. The Fraunhofer IPK and arvato services specialists with the help of expert interviews will define comparison criteria for technological approaches that attempt to satisfy the demanding requirements of a digitisation process. This is meant to find the optimal implementation of adequate technology. The framework for this phase is provided by the information gained through the characterisation of the IAI library, the conducted survey and the collected experiences of practical users of digitisation techniques. Furthermore, Fraunhofer IPK and arvato services will put together their knowledge about virtual reconstruction of destroyed documents to complement this assessment process.

Digitisation offers many benefits for our cultural heritage as well as our libraries and their users. As it turns out, this implies a wide range of requirements and an evaluation of which technical solutions are needed in order to fulfil them. After having collected all demands, our study will conceptualise a solution that is beneficial and flexible in terms of technology and workflow, and which puts us in the position to say that when it comes to digitising printed cultural heritage, we are not caught between user demands and preservation.

References:

[1] Praxisregeln Digitalisierung by the German Research Foundation (DFG):
[2] http://www.dfg.de/forschungsfoerderung/wissenschaftliche_infrastruktur/lis/download/praxisregeln_digitalisierung.pdf

Abstract
As the whole field of preserving, documenting and studying the Cultural Heritage is interdisciplinary, and the way in which information is managed is not homogenous. Moreover the objects belong often to the real world and present a 3D component, not easy to represent using only two-dimensional approach. Storage, organisation and retrieval of such information in real time is challenging and commonly not very well structured. Very often the only unifying entity is the "object", which the information is related to, so an effective management of related data still represents a serious problem. 3D visualisation simulates spatial reality, allowing the viewers to more quickly recognise and understand what they see in the real world. Cultural heritage draws together several different professions. Furthermore, the relationship between the conservation managers, who are often unfamiliar with current documentation techniques, and the providers of the information, who tend to be highly technical practitioners without expertise in cultural heritage, is not easy to handle. We present a new method to access spatial information through the interactive navigation of a synthetic 3D model, which reproduces the main features of a corresponding real environment. The information is ranked with a novel measure of the relevance, that depends on the position/orientation in the 3D space, allowing users to retrieve significant information. To give access to a larger audience, the method is accessible through an intuitive and user-friendly interface on normal Web browsers.
The system has been applied to case studies related both to outdoor and indoor environments. Actually the developments are relative as an interactive smart guide.
In particular we believe that an intuitive interaction in real time and in the context makes more accessible the information, and can help users in being more active, and learn in interesting ways.
Keywords: interactive 3D interface; relational database; gaussian; spatial relevance; overlap; XVR 3D engine.

Introduction

The Cultural Heritage normally refers to objects in the real world with a 3D component and often requires a uniform treatment to massive heterogeneous data. To keep in account these issues, the research started to focus on 3D representations. Nevertheless, using a 3D environment allows for a closer adherence to the real world (preserving location related data) and permits to respect the spatial relationships among different components. Our aim has been to develop a new approach to access and manage information, paying particular attention to cultural assets data management. This approach, called ISEE ("I see"), will be based on "interactive 3D models", because an interactive interface allows for a more natural behaviour, where the user can move freely and find the sections he/she is interested in. ISEE should be able to provide retrieving information by just looking around in a 3D environment, as moving and looking at the world is the main modality we use to gather information from it.

State of art

Today, in the field of Cultural Assets, it is very important to continuously research new ways to represent and query data. Normally one has to deal with information from different sources and formats, and with a lot of data produced in a short amount of time. Another problem is the communication between who provides recording, documentation and information management tools, and the professionals in cultural heritage management who use them. The ICOMOS/ISPRS Committee for Documentation of Cultural Heritage (CIPA Heritage Documentation) conducted a series of workshops between 1995 and 1999 to understand this incompatibility. Often conservation managers are unfamiliar with current documentation techniques, while the providers tend to be highly technical practitioners without expertise in cultural heritage [1]. One possibility to communicate the structure of a piece of information is to visualize it using a graphical representation. "Information visualization" is a wide field interdisciplinary in nature and represents one important process to transform and represent a large variety of data.

Chiara Cirinnà and Maurizio Lunghi (edited by), *Cultural Heritage on line: Empowering users: an active role for user communities. Florence, 15th - 16th December 2009* ISBN 978-88-6453-184-7 (print) ISBN 978-88-6453-187-8 (online)

Internet has changed the way we organise data, but an integrated management method for cultural heritage ICT applications is still not available. The Virtual Reality (VR) or the Augmented Reality (AR) technologies are able to reconstruct 3D models of ancient culture, making them accessible to modern-day users [2]. This can be useful for members of the general public, but for specialists working in the fields of archaeology or restoration, this approach is not necessarily so useful and can even be misleading. 3D interaction is an intuitive paradigm for a majority of people and additionally can convey more information. It is clear that the problem of moving from 2D to 3D is complex. The use of virtual reality can help to understand and to manage the real world, but the transition between these realities is not easy and not always the result satisfies the expectations. Often one can see a 3D model with a high level of detail and it's an excellent work. But it's not always possible to read the information while navigating within the model. It's necessary to decide "what must have the priority". Actually, Web applications as ISEE are increasingly used, because they allow access from any computer, while keeping a centralised repository of information. The quality of their interface has practically reached parity with desktop applications.

Motivation

We have chosen an interactive interface, because we want to involve the user: interacting one can learn more. We think that an intuitive interaction, such as looking at the environment where information is contextualised, is one of the solutions closest to normal human behaviour. Our intention is to create an environment that enables the co-operation and the exchange of knowledge among users. The interface, we propose for ISEE, enables the user to explore a three-dimensional space, where the objects are geographically referenced, and retrieve the related information. In particular professionals from the field should benefit from a greater level of freedom to manage and manipulate the information retrieved, with tools specific to their field. They will be able to insert more detailed data and to decide presentation and retrieval modes.

The method

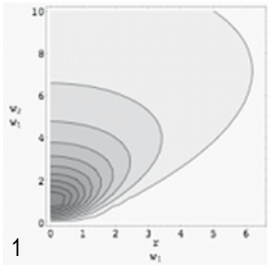
Our approach started considering the requirements for an adaptive and intuitive interface to access information. Thinking about it we had the idea of using the simple action of seeing as "a common language" to query and insert information. To achieve this, one has to define for each "view" the region on which the user is focused in that moment, that we called View Zone (VZ). At this point one can either be interested to recover the relevant information about that zone, or add new information about it. The complexity dictated by the type of data is simplified in a few easy moves (navigating an environment and looking around).

We decided to treat approximated zones to represent the objects in 3D space. In the method our information is associated with regions of the space, which we called Information Zones (IZs). The Information Zone does not have to coincide with a 3D object represented in the model, but they might be just a part of it or include many objects at the same time.

To define the region in a precise way that a computer can understand we used "3D gaussians": this is a function which assigns a value to each point of the space (it can be seen as a fog, more dense in the centre, and less dense on the periphery), which can be used to adequately describe the concentration of information. The interactive 3D viewer (VZ) and the IZ are approximated with a normalized 3D gaussians, and this provides to have a symmetric treatment. It allows us to use the interactive 3D viewer to visually insert the IZ of a piece of information (authoring), or to jump immediately to the view related to some information (retrieval). An innovative aspect is the definition of spatial relevance of information. A ranking calculates our relevant information and depends from the View Zone and the location of the information. The measure of relevance is depending on the spatial relationships between VZ and IZ (Figure 1). Intuitively, the relevance of information should be "maximal" (relevant) when its Information Zone (IZ) coincides with the View Zone (VZ) (Figure 2), decreasing when they are far apart. An IZ that has a size comparable to the current VZ is probably more interesting than an IZ that has a size very different from the current VZ.

Using as technique of accessing information our overlap, based on the distance and the size of the VZs and IZs, is more accurate respect of other systems based on distance (e.g. GoogleMaps) or on the selection (3D interfaces, games etc.). These methods typically use the distance from the user not from where he/she is looking at as extra simplification. In an interactive 3D model "the level of zoom" depends on the distance of the object looked at, and in the same scene one can have different levels of detail.

Moreover the system can manage with high densities of data, because its usage provides an extra means to filter the information reduces the amount of information retrieved.



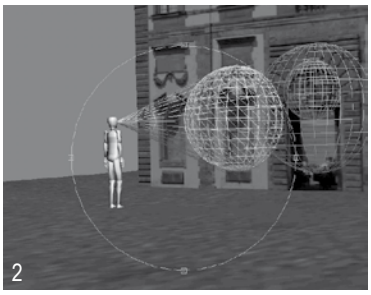


Figure 1 Contour plot of the overlap of two normalized gaussians with width "w1" and "w2" and a distance "r" between the two centres. Darker means a higher overlap.
Figure 2 Third person view of VZ and IZ from the side, the VZ coincides with an IZ and has maximum overlap. The VZ is represented as a green sphere, the IZ as red spheres, and the view cone is gray.

The case studies

The prototypes developed in this work represent Web applications, where the user can explore in intuitive way a model and to discover the information linked to it. In order to visualize and interactively navigate the model on the Web, we used the XVR technology [3], jointly developed by PERCRO Scuola Superiore Sant'Anna in Pisa (Italy) and VRMedia s.r.l.. The 3D model is downloaded on the user client as soon as the user accesses the Web page. As soon as the download is finished, a first list of information automatically appears, presenting on the top the data most relevant for the zone the user is currently looking at. The structure of the archive implemented so far is quite simple. The information is registered in a file system (in xml, jpeg, tiff etc.) and stored as meta-information in the relational database (MYSQL) to have a fast access in real time (query, add etc.). A nice consequence of storing Il information also as file is that normal tools for the automatic indexing of files could be used in the future to index the meta-information and document files to allow full text searches [4] even if they do not take into account temporal and spatial information. The last version of ISEE can upload files in kml, a format standard of Google Maps.

The development of the method started from a first case study: the crypt in St. Servatius in Quedlinburg (Sachsen-Anhalt, Germany), part of the World Heritage List (UNESCO). It presented different types of information related to the restoration and a cloud of points from a 3D scanner. The crypt represents one of the largest painting cycles of the 12th century in Germany [5][6][7]. The research and the work developed by Prof. H. Leitner and his students of the Hochschule für Bildende Künst of Dresden and is work in progress too. In particular it has been realized using GIS format extensively as a documentation tool, with the "base map" consisting of high-resolution rectified georeferenced photographs. The crypt can represent a prototype for sharing, query and add the information among professionals figures or common users (Figure 3).

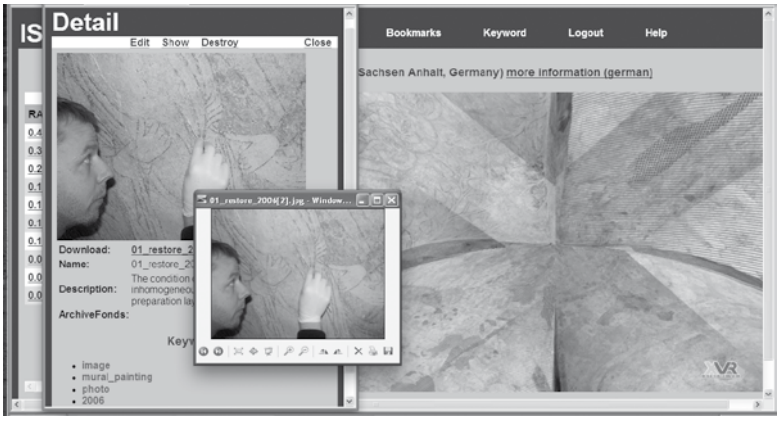


Figure 3 The interface of the crypt of St. Servatius: the mural paintings and the maps of the state of conservation in the same view.

DEVELOPMENTS

We are working for a use of a Web application supporting all browsers and to provide ISEE as city sightseeing. In the past the method had been already applied in real world with good results, using a GPS Compass (Vector CSI Wireless), providing 2D heading and positioning data connected to a laptop in Piazza Napoleone in Lucca, Italy [8]. The actual system on smart device works by a similar approach of the web application. The data itself is gathered and stored using a REST interface to the ISEE Web server (Figure 4a-4b). For efficiency reasons the interaction on the device is mostly 2D.

Figure 4 a) The actual development on mobile device b) The application ISEE in Berlin

CONCLUSIONS

The interactive visualization makes information more accessible and improves the user experience. The work presented provides an intuitive and user-friendly interaction for accessing, inserting and modifying information in a 3D space. The method is suitable to all categories of users, both professional and non professional, because it is based on the simple action of navigating the 3D space and retrieving the information. Moreover the information is associated with regions of the space, and thus pre-processing of 3D models to subdivide them in suitable logical elements is not needed. New pieces of information can be inserted in the same way in which they are queried, just by looking. The use of extended zones allows us to use a ranking algorithm with superior performance than rankings based only on the distance. The proposed ranking algorithm matched the intuitive expectation of the users, as was verified with a formal usability test that was performed at completion of the work. The method we propose is intended to: allow easier information handling; use only simple and standard formats, in order to facilitate communication and exchange; provide the option of detailing the information source.

REFERENCES

[1] http://www.getty.edu/conservation/field_projects/recordim/index.html
[2] B. A. Doug, E. Krujff, J. La Viola and I. Poupyrev, 3D User Interfaces, Theory and Practice, Addison-Wesly, USA, 2005.
[3] M. Carrozzino, F. Tecchia, S. Bacinelli, and M. Bergamasco, Lowering the Development Time of Multimodal Interactive Application: the Real-life Experience of the XVR project, In: Proceedings of ACM SIGCHI International Conference ACE, November 19 – 22, Valencia, Spain, 2005, pp. 270-273.
[4] L. Pecchioli, F. Mohamed, A method to access the information through an interactive 3D virtual environment, In: Entwicklerforum Geoinformationstechnik- Junge Wissenschaftler forschen – Technische Universität Berlin, Institut für Geodäsie und Geoinformationstechnik, Berlin, Deutschland, (Eds) Shaker Verlag, 2008, pp.119 – 129.
[5] Gosslau, Friedemann and Radecke, Rosemarie, Die Stiftskirche zu Quedlinburg, Eine Fürung durch den romanischen Sakralbau und den Domschatz, (Eds.) Convent-Verlag, Quedlinburg, 1992.
[6] H. Leitner, La conservazione delle pitture murali nella Cripta di San Servatii a Quedlinburg, In Arcadia Ricerche Eds., Proceedings of Bressanone, Scienza e Beni Culturali, XXI - Sulle pitture murali, Bressanone, Italy, 12-15 July 2005, pp. 233-240.
[7] L. Pecchioli, F. Mohamed, M. Carrozzino, ISEE: accessing information navigating in a 3D virtual Environment. The case study of the crypt in St. Servatius in Quedlinburg, Saxony-Anhalt (Germany), In: Web Portal - Architectural Image-Based-Modeling, 2009.
[8] L. Pecchioli, F. Mohamed, M. Carrozzino, H. Leitner, Accessing information through a 3D interactive environment, In: Proceedings of ICHIM07 Digital Cultural and Heritage, Toronto, Ontario, Canada, 2007. (http://www.archimuse.com/ichim07/papers/pecchioli/pecchioli.html)

TUTORIAL
LONG TERM PRESERVATION OF DIGITAL ASSETS:
BASIC CONCEPTS AND PRACTICES
Monday 14th December

The event brought together international experts who developed a one-day full immersion tutorial about issues related to long term preservation of digital objects. The tutorial started setting the scene about current initiatives and approaches, then it gave participants an understanding of the key digital preservation issues and decisions to be taken during the lifelong cycle of a digital archive. Some clear concepts, recommendations and "to do" list of things were presented. The major challenges and the most prospective solutions were introduced, even if findings in this field are not so mature. The experts defined needs and experiences about the specific cultural heritage sector, providing the audience with some technical recommendations about standards on digital archives, metadata, digital formats, strategies and criteria to certify tools and practices, check risks and help to take the right decisions for preservation planning. After lunch, the session started with an interactive hands-on work, where concrete experiences and practical tools developed by some of the most important European projects were presented and demonstrated. Target audience were librarians, archivists, museum curators, students, researchers and professionals in the sector of digital archives management, digital libraries, Internet applications and multimedia content creators.

TUTORIAL
DUBLIN CORE - BUILDING BLOCKS FOR INTEROPERABILITY
Thursday 17th December

This Tutorial describes the Dublin Core Metadata Initiative, the history of the organization from a group of interested experts in 1995 to a formal organization with the legal incorporation in late 2008, and outlines the strategic directions and collaboration with other metadata initiatives over the years and the strategic directions that DCMI will be pursuing in the near future. After a brief outline of the organizational history of DCMI with presentation of the communities, task groups, processes, committees and operating rules, the tutorial will provide a general introduction to Dublin Core metadata, technical trends in the Dublin Core community over the past decade, and alternative approaches to descriptive metadata in the "Dublin Core" style. The second part of the tutorial reviews implementation technology alternatives using HTML, XML, and RDF, including new techniques such as embedded RDF a metadata in support of structured search. Alongside the traditional paradigm of metadata interoperability on the basis of pre-coordinated agreements on natural-language definitions and specific data structures, the new paradigm of Linked Data offers a flexible framework for coherently merging diverse types of metadata on the basis of a shared underlying data model. The tutorial covers methods for expressing controlled vocabularies as Linked Data such as Simple Knowledge Organization System (SKOS).

Chiara Cirinnà and Maurizio Lunghi (edited by), *Cultural Heritage on line: Empowering users: an active role for user communities. Florence, 15th - 16th December 2009* ISBN 978-88-6453-184-7 (print) ISBN 978-88-6453-187-8 (online)

Finito di stampare da Grafiche San Benedetto srl
Castrocielo (Fr) - ITALY