DE GRUYTER
SAUR

Dominic J. Farace,
Joachim Schöpfel (Eds.)

GREY LITERATURE IN LIBRARY AND INFORMATION STUDIES

DE
G

Grey Literature in Library and Information Studies

Grey Literature in Library and Information Studies

Edited by
Dominic J. Farace and Joachim Schöpfel

De Gruyter Saur

ISBN 978-3-598-11793-0
e-ISBN 978-3-598-44149-3

Library of Congress Cataloging-in-Publication Data

Grey literature in library and information studies / edited by Dominic J. Farace and Joachim Schöpfel.
p. cm.
Includes bibliographical references and index.
ISBN 978-3-598-11793-0 (acid-free paper)
1. Grey literature. 2. Grey literature--Bibliography--Methodology. I. Farace, Dominic John. II. Schöpfel, Joachim, 1957-
Z1033.G73G74 2010
025.2'82--dc22

2010029505

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at http://dnb.d-nb.de.

Typesetting: Michael Peschke, Berlin
Printing: Hubert & Co. GmbH & Co. KG, Göttingen
∞ Printed on acid-free paper

Printed in Germany

www.degruyter.com

Contents

Part II – Uses, Applications, and Trends in Grey Literature

Introduction Grey Literature

Dominic J. Farace, GreyNet International, Netherlands
Joachim Schöpfel, University of Lille, France

0.1 Definitions

Knowledge generation in any field of studies begins with clear, accepted or at least conventional definitions of terms. Through the years, a number of uncontrolled terms have been used to describe the phenomenon of grey literature. This has not really contributed to the understanding, use, and application of grey literature. In 1997, the definition of grey literature often referred to as the 'Luxembourg Convention' took a sharp turn – emphasizing for the first time the supply side of grey literature, that is its production and publication both in print and electronic formats. This break from the previous quarter century, which narrowly focused on the demand side and the problems of bibliographic control, indexing, cataloging and retrieval finally placed grey literature in its fuller perspective.

The definition of grey (or gray) literature accepted during the Third International Conference on Grey Literature in Luxembourg reads " … that which is produced on all levels of government, academics, business and industry in print and electronic formats, but which is not controlled by commercial publishers"[1]. During the Sixth International Conference on Grey literature in New York City, a postscript was recommended to that definition and shortly thereafter added: " i.e. where publishing is not the primary activity of the producing body".[2]

Another definition is from the U.S. Interagency Gray Literature Working Group, "Gray Information Functional Plan," 18 January 1995, which defines grey literature as "foreign or domestic open source material that usually is available through specialized channels and may not enter normal channels or systems of

1 Farace, D.J. (1998), Foreword - In: Third International Conference on Grey Literature : Perspectives on the Design and Transfer of Scientific and Technical Information, 13-14 November 1997 in Luxembourg. GL'97 Conference Proceedings, p. iii. - (GL Conference Series, ISSN 1386-2316 ; No. 3). ISBN 90-74854-17-6

2 Schöpfel, J., C. Stock, D.J. Farace, and J. Frantzen (2005), Citation Analysis in Grey Literature: Stakeholders in the Grey Circuit. – In: The Grey Journal, vol. 1, no. 1, pp. 31-40. – ISSN 1574-1796.

publication, distribution, bibliographic control, or acquisition by booksellers or subscription agents".[3]

In fact, the term traditionally covers three categories of documents – conference proceedings, reports and doctoral theses – often printed in small numbers. Nevertheless, the borderline with "white" or "conventional" literature is permeable, since some conference proceedings are published by commercial publishers as monographs or in serial publications such as journals. The same holds true for some reports. Likewise for doctoral theses, especially in the humanities and social sciences, some are found on the commercial publishing market.

However, regarding the multitude of other documents that circulate outside conventional publishing, the lack of "commercial control" raises real problems for academics and scientists as well as for information professionals when it comes to locating and acquiring them. The lack of "commercial control" and promotion also often implies a lack of "bibliographic control". In other words, these documents are often inadequately referenced in catalogues and databases, so that searches through this category of scientific information require specialized knowledge on sources and grey circuits.

0.2 A short history

Library and information professionals have been contributing to studies on grey literature for nearly 30 years now, compiling a rich corpus of articles, reports and conference papers.

The Grey Journal from TextRelease/GreyNet in Amsterdam, the only current journal dedicated to this topic, published some 100 articles since 2005. Another serial, *The International Journal on Grey Literature*, was edited by Emerald (former MCB University Press) but ceased publication in 2001. Most other articles on grey literature are published in serials in library and information sciences or journals from other scientific domains such as *The Lancet*, *Marine Policy*, or *European Psychiatry*. And to date, only one other monograph has been published on grey literature.[4]

Since 1992, the Grey Literature Network Service (GreyNet) organizes international conferences on grey literature that have already taken place in Amsterdam, Netherlands (1993, 2003 and 2008), Washington D.C. (1995, 1999 and 2009), Luxembourg (1997), New York City (2004), Nancy, France (2005), New Orleans, Louisiana (2006), Antwerp, Belgium (2007) and Prague, Czech Republic (2010, Forthcoming).

---

3 The U.S. Interagency Gray Literature Working Group definition of grey literature, http://en.wikipedia.org/wiki/Gray_literature

4 Auger, C.P. (1998) Information Sources in Grey Literature. 4th edition. – London : Bowker-Saur, 177 p. – ISBN 1-85739-194-2.

The more than 250 authors and researchers in the field of grey literature, who have contributed to the above conference programs form as it were the WHOIS in Grey Literature along with the host and sponsoring organizations, whose financial contributions guarantee the continuity and longevity of research programs and projects in the various sectors of government, academics, business and industry.

The TextRelease website provides biographical notes for over 75 academics, scientists and professionals who work and publish in the field of grey literature.

Five outstanding personalities made lasting contributions to specific areas in the field of grey literature in the four decades from 1960 to 2000: Alvin M. Weinberg (United States) author of the famous "Weinberg Report", Vilma Alberani (Italy) organizer of a national program for grey literature, Charles P. Auger (United Kingdom) who provided the first Roadmap of Grey Literature Systems and Services, Ulrich Wattenberg from the German Max-Planck-Gesellschaft who specialized in the infrastructure of grey literature for the Japanese scientific and technical information, and Andrei Zemskov (Russia) from VNTIC, the National Public Library for Science and Technology, where he explored the free access of information and grey literature.[5]

We can distinguish five periods for the development of research and the development on grey literature.

1. They begin with the years leading up to 1979 in which numerous uncontrolled terms such as ephemera, fringe literature, fugitive literature, non-conventional literature, non-published literature, report literature, research outputs, small-circulation literature, unconventional literature, unpublished literature, etcetera were coined to capture the growing phenomenon.
2. The period 1980-1990 covered the development and launch of national and international programs on grey literature (1985 is the year in which the European network EAGLE was created).
3. 1990-2000 included the creation of GreyNet, the Grey Literature Network Service (1993 is the year in which the first international conference on grey literature was convened).
4. The years 2003-2005 covered the re-launch of the Grey Literature Network Service showcasing new projects in the context of the explosion of digital resources, the movement for open access to scientific and technical information, and the Web2.0 (these research results were presented at GL conferences in Amsterdam 2003, New York 2004, and Nancy 2005. This growth occurred notwithstanding the fact that EAGLE and its SIGLE database (System for Information on Grey Literature in Europe) was discontinued in 2005.

5 Farace D.J. and J. Frantzen (2004) Four winds on the grey landscape: a review of four information professionals, their work and impact on the field of grey literature. – In: Fifth International Conference on Grey Literature: Grey Matters in the World of Networked Information Amsterdam, Netherlands, December 4-5, 2003. GL5 Conference Proceedings, pp. 10-12. (GL Conference Series, ISSN 1386-2316 ; No. 5). ISBN 90-77484-01-9

5. The current timeframe from 2006 onward is one in which new cooperative research initiatives in the aftermath of EAGLE-SIGLE are on the rise.

One of the recent projects is the OpenSIGLE project, an initiative powered by INIST (France) to provide access to former SIGLE records in an open source context.[6] In the spring of 2008, GreyNet signed on to the OpenSIGLE Repository in order to preserve and make openly available research results originating in the International Conference Series on Grey Literature. And, in so doing, the OpenSIGLE Repository has become the intersection of more than 25 years of bibliographic information on grey literature with 15 years of research in the field.

Another initiative is the collaboration of researchers in the field of grey literature on institutional levels involving cross-country and international partnerships. And yet another recent initiative was the pilot for a distance learning course on grey literature for (post)graduate students, one that was accredited by the University of New Orleans (UNO) and which is now available to other academic institutions.

0.3 Typology

We indicated earlier that the term grey literature traditionally refers to reports, conference proceedings and doctoral theses.

Reports are the most numerous by far among the different types of grey literature in the OpenSIGLE database. But the 'reports' category covers a wide variety of very different documents: institutional reports, annual or activity reports, project or progress reports, technical reports, reports published by ministries, laboratories or research teams, etc. Some are disseminated by national and international public bodies. Others are confidential, protected, or disseminated to a restricted readership, such as technical reports from industrial laboratories. Some are voluminous, with statistical appendices, while others are only a few pages in length.

In the other categories, citation analyses offer a tremendous range of grey resources.[7] Besides theses and conference proceedings, they also include unpublished manuscripts, newsletters, recommendations and standards, patents, technical notes, product catalogs, data and statistics, presentations, personal communications, working papers, house journals, laboratory research books, preprints, academic courseware, lecture notes, and so on. GreyNet in fact maintains an extensive online listing of document types, which are categorized as grey literature.

---

6 See chapter 9 in this monograph.

7 Farace, D.J., J. Frantzen, J. Schöpfel, C. Stock, and A.K. Boekhorst (2006) Access to Grey Content: An Analysis of Grey Literature Based on Citation and Survey Data : A Follow-up Study. – In: Seventh International Conference on Grey Literature: Open Access to Grey Resources, Nancy, France, December 5-6, 2005. - GL7 Conference Proceedings, pp. 194-203. - (GL Conference Series, ISSN 1386-2316 ; No. 7). ISBN 90-77484-06-X

However diverse, these documents all share one thing in common, they contain unique and significant scientific and technical information that is often never published elsewhere. The lack of descriptive referencing and adequate circulation is therefore, as we have said earlier, a real problem for scientific communication.

The Internet, however, is now altering the entire landscape. Not only because of changing user behavior, but also, and especially, because more and more grey literature is being published on the Web. As one study from the German Centre for Information in the Social Sciences has pointed out, the switch from paper to digital does not necessarily mean that more grey literature is appearing.[8] Instead, the Internet has radically changed access and distribution methods, accentuating the ephemeral and volatile nature of grey literature. This same study also drew attention to the fact that many journals and the journal articles contained therein can be categorized as grey literature i.e. where publishing is not the primary activity of the producing body. The fact that in Europe, for more than two decades the SIGLE database did not identify journals and journal articles as grey literature may account in part for the apparent neglect of these two types of grey documents.

And yet, another special type of grey material is also likely to gain more importance. Until now, raw data – the basis for many scientific publications – are widely unpublished and inaccessible. Today, public research organizations are starting to develop national and international strategies for the control and archiving of these files, the data, and statistics.

0.4 Challenges

Grey literature will remain a challenge for information and documentation professionals as well as an interesting field for research activities in at least six areas:

The need for a new definition: The traditional definition of grey literature needs to be further refined and/or redefined by way of an accurate analysis of new means of access and distribution, in line with Mackenzie Owen's observation that "Grey does not imply any qualification (but) is merely a characterization of the distribution mode".[9] What we see is that the current 'Luxembourg' definition moved from emphasis on the acquisition of grey literature to the production of grey literature. And now, the definition should reflect both.

The need for a new 'value chain': In the Netherlands, Roosendaal has in the past few years, been examining the process whereby universities re-appropriate publications. In his work, he highlights the radical changes taking place in the

8 Artus, H.M. (2005) Old WWWine in New Bottles? Developments in electronic information and communication: structural change and functional inertia. *The Grey Journal*, vol. 1, no. 1, p. 9-16. – ISSN 1574-1796.

9 Mackenzie Owen J.S. (1997) The Expanding Horizon of Grey Literature. In *Third International Conference on Grey Literature: Perspectives on the Design and Transfer of Scientific and Technical Information.* GL3 Conference Proceedings, Luxemburg, Nov 13-14, 1997; Commission of the European Communities DGTIMER, Luxemburg.

'value chain' of scientific publication.[10] This type of research and evaluation of scientific publications brings to the forefront major issues in the context of emerging STI trends. What is the future of peer review? Which "quality label" applies to working papers or scientific communications on blogs or in open repositories? Does the community approach of Web 2.0 offer a viable solution for the need for quality standards of non-commercial STI materials? The impact of new technologies in information and communication on the dissemination of non-conventional literature is a complex matter and the potential field for research is vast. To date, research and analyses have only broken ground giving way to a vast and virtually untapped field of investigation.

The need for an economic model: Collecting, distributing and searching grey literature all come at a price, which may in fact be much higher than for journal article and book searches. To date, there is no clear economic model in this area and further analysis is needed in terms of investments, direct and indirect costs, acquisition prices, and the like. The case of EAGLE underlines the need for public funding and a sustainable economic model to guarantee the bibliographic coverage as well as full-text, enriched dissemination of grey literature.

The need to oversee archiving practices: New technologies for information and communication facilitate resource archiving in general, and there is strong incentives from the "open access" movement. Nevertheless, the question of "who should archive what, where, when, and for how long" has remained largely unanswered. Aware of information policy and the concomitant financial aspects involved, answers are rather urgently needed, even if they now were only able to address part of grey literature resources.

The need to clarify the legal aspects: The legal status of grey resources and rights in their use (deposit, archiving, distribution, etc.) is a major challenge for the future of this form of STI publishing. The national and international legal environment is evolving rapidly, and all restrictions, exceptions and technical constraints (e.g. digital rights management, interoperability etc.) of the new laws on intellectual property, author's rights and copyright also apply to grey resources. Nevertheless, very few documentary analyses have addressed legal aspects in the field of grey literature and their subsequent economic consequences.[11]

The need for education and training: Over the past years, training courses, guest lectures, seminars and workshops have been organized by information professionals on the topic of grey literature. Most of these endeavours have undoubtedly had some impact on this field of information. As mentioned earlier in the chapter, an accredited college course on grey literature is carried out via the University of New Orleans' (UNO) distance education program since 2007. Education and training is fundamental to the future of grey literature - not only for LIS students and their instructors but also for information professionals and practitioners in government, as well as business and industry.

10 See chapter 1 in this monograph.

11 See chapter 6 in this monograph.

0.5 Further Considerations

In concluding our introduction to this monograph on grey literature, we offer the reader still other prospects in need of further reflection. And, we are confident that they will be duly addressed in the subsequent chapters in this book.

It seems likely that

Grey literature will not disappear, but will continue to play a significant role alongside commercial publishing. Our research has led us to believe that information discovery into the various types of grey literature available in print and electronic formats is ever increasing.

The borderline between "grey" and "white" (commercial) literature will become increasingly indistinct, particularly in an environment that is moving towards open access to STI.

The proportion of "grey" documents published on the Web will continue to increase. We see this development closely linked to the production of grey literature in digital environments, as well as to retrospective activities leading to republication.

The Internet will encourage a greater diversity in the types of "grey" resources available such as (raw data, personal notes and comments, lectures, newsletters, product catalogues, etc.).

It also seems likely that

Bibliographic control of grey literature will remain problematic despite the trend towards standardization of digital documents. We find that this has everything to do with the application and use of standards, which are in transition.

Open archives will offer more appropriate services and functions for at least some segments of grey literature i.e. preprints, doctoral theses, and reports. We mention these three types of grey literature, because they have come to form special collections making them more visible in and for repositories.

Some organizations – especially in the public sector (e.g. national libraries and STI centers) but also in the private sector (e.g. Elsevier, Google, etc.) – will develop tools and services to aid in the efficient exploitation of grey resources on the Web. This in all likelihood is based on the response by such organizations to research efforts by the global grey literature community.

However, it seems unlikely that

Searching and collecting grey literature will become as straightforward as it is for journals and books in the traditional publishing sector. We adjudge that the increase in grey over commercial publications is the main explanation for this.

New tools for collecting, depositing, and archiving will make grey literature less ephemeral and volatile than in the past. Our research indicates that until an organization formulates a policy on grey literature backed by budget appropriations, the implementation of technology cannot be guaranteed and thus the environment in which grey literature has coexisted in the past will remain unstable in the likely future.

Part I, Section One

Producing and Publishing Grey Literature

"Grey does not imply any qualification (but) is merely a characterization of the distribution mode"[1]. The current 'Luxembourg' definition moved from emphasis on the acquisition to the production of grey literature. The first section in this book looks at three studies on the production and publishing of grey literature in the field of scientific and technical information written by academicians in economics, library and information sciences.

In the Netherlands, Roosendaal among others has examined the process whereby universities re-appropriate publications. He highlights the radical changes in the value chain of scientific publication triggered by the potential that information and communication technology offers the author and reader. His chapter revisits work carried out in 2003, emphasizing new business models for scientific publishing.

One of the conclusions is that "research and higher education institutions are the natural candidates to initiate the development of new business models and structures. This is foremost an organisational and not a technical challenge. A major organisational challenge will be to absorb the library consequently into the research organisation."

The second chapter, 'How to assure the quality of grey literature, the case of evaluation reports' is in essence Weber's study on the quality assurance system by the Swiss Federal Office of Public Health. 'Report quality' is defined by the quality of processes, tools, and conduct applied throughout the study. The study does not claim a universal system for all producers and types of grey documents but considers that "a basic set of steps for guiding the production of quality output" could improve the overall quality of grey literature. Could such a system be generalised? Well, the recent debate on quality and reliability of grey research reports[2] gives emphasis to the relevance and actuality of this analysis.

The final chapter of this section offers an overview of the production and processing of another category of grey literature. Južnič from the University of Ljubljana in Slovenia draws on experiences, initiatives, and projects from different

---

1 Mackenzie Owen J.S. (1997) The Expanding Horizon of Grey Literature. In Third International Conference on Grey Literature: Perspectives on the Design and Transfer of Scientific and Technical Information. GL3 Conference Proceedings, Luxemburg, Nov 13-14, 1997; Commission of the European Communities DGTIMER, Luxemburg.

2 ClimateGate: the mistake on glacier melting introduced in the 2007 UN Intergovernmental Panel on Climate Change (IPCC) report.

countries: United Kingdom, France, Slovenia, India, South Korea, etc. and develops a framework for electronic theses and dissertations. Južnič anticipates that "it will be exciting to see (…) grey literature (…) become the core of higher education activities and a centrepiece of a university's reputation."

In the compilation and editing these three chapters, it was *not* our intention to provide a coherent and exhaustive economic or social theory on the production and publishing of grey literature. Rather instead, to suggest to the readership that they keep in mind certain key questions drawn from the authors' works, namely:

What is the specific function of grey literature in the communication process of scientific communities? How does the Internet impact this function? What is or could be the role of academic libraries in the production and publishing of grey literature? And, how does one guarantee an acceptable level of quality for grey documents?

Chapter 1

Grey Publishing and the Information Market: A New Look at Value Chains and Business Models

Hans E. Roosendaal
University of Twente, The Netherlands

Justification

The article "The Information Market for Research and Higher Education"[1] was written on the occasion of the Fifth International Conference on Grey Literature held December 4-5, 2003 in Amsterdam.

Since then, the author has been involved in a number of publications (Roosendaal et al., 2005[2]; Roosendaal et al., 2008[3]; Roosendaal et al., 2009[4]) further developing the subject of the article albeit not strictly focusing on grey literature. In particular, the last two publications, a book chapter and a comprehensive book are recent and report new developments.

In this article, the author has chosen to make use of the 2003 article in combination with Roosendaal et al., (2009) with a focus on aspects of grey literature. As main source, Roosendaal et al., (2009) will be briefly but comprehensively quoted without mentioning this explicitly. For further details on the discussed issues, the reader is advised to consult Roosendaal et al. (2008, 2009).

The parts of the article that are copied from the 2003 article are taken over verbatim and are recognisable as printed in *italics*.

---

1 Roosendaal H.E., (2004) "The Information Market for Research and Higher Education, *How to integrate all relevant information in a network of repositories?*" Publishing Research Quarterly, 20 (1), p. 42-53.

2 Roosendaal H.E., P.A.Th.M, Hilf E.R, (2005) 'Pertinent Strategy Issues in Scientific Information and Communication in 2004', Invited review in *Library Science- quo vadis?*, edited by Petra Hauke, Institute of Library Science at the Humboldt University Berlin, Berlin, K.G. Saur Verlag, München, pp: 217- 238.

3 Roosendaal H.E., Kurek K., Geurts P.A.Th.M. (2008). 'Modèles économiques de l'édition scientifique et processus de recherche' in J. Schöpfel, *La publication scientifique. Analyses et perspectives*. Hermes Science, Lavoisier.

4 Roosendaal H.E., Zalewska-Kurek K., Geurts P.A.Th.M. Hilf E.E. (2009) *Scientific Publishing, from Vanity to Strategy*. Chandos, Oxford.

1.1 Introduction

"Authors want to publish more, readers want to read less." This statement paraphrases the fact that wide exposure is paramount to the author and (pre)selection to the reader of research information, including grey information. Any force in the market like the use of Information and Communication Technology (ICT) by the actors involved (authors, readers, libraries, scientific publishers etc.) that allows better fulfilling this statement is an engine for change in the value chain, prompting changes in the roles of the stakeholders in scientific communication.[5]

The above statement means that, for the author, visibility is crucial whilst, for the reader, retrievability is. In this context it is important to bear in mind that readers, when searching for information, will in most cases not be able to specify in detail what they are looking for. Combining these various factors can only lead to the conclusion that wide availability of information is the foremost requirement in this market. Arguing along the familiar business criteria of volume and margin we see that wide availability takes the role of high volume and restricted availability that of low volume. In the research and higher education (HE) information market volume is thus the potential volume of readers, rather than the actual volume of reading. The fact that readers want to read less but everything that is relevant to them at the right time illustrates this point of view. This means that the elasticity in the market is determined by the degree of availability, and this is compatible with the requirements for an open system.

This discussion illustrates that the statement at the head of this introduction determines to a large extent the dynamics of the market, and is independent of the carrier of the information, be this paper or a digital carrier. In other words, the value chain of the research and HE information market is largely determined by it. In this value chain the author and the reader, jointly the user, are the generic stakeholders while other stakeholders are institutional stakeholders.

The main driving force in the market is thus seen to be the desire of researchers to share information with the research community and the wider societal community. E-science can be seen as a further step towards the ideal of universal sharing of scientific results and making research information an ever more integral part of the research process. E-science is an integrative concept: it comprises not only the changes in the process of sharing information but also and above all new opportunities in the research process itself.

The gist is that e-science is a further step in making research information the integral raw material in the research process as it should be. In e-science, it will be possible to share primary data much more efficiently with other researchers allowing for new schemes of division of labour e.g. in splitting up collecting data in an

5 Roosendaal H.E., Geurts P.A.Th.M., van der Vet P.E. (2001) Developments in scientific communication: Considerations on the value chain. *Information Services & Use*, vol. 21, p. 13-32.

advanced way from analysing these same data and so on, as is daily practice in e.g. high energy physics.

E-science thus leads to new research strategies and research communication strategies with the goal to improve the production of new knowledge. Researchers will have to develop clear strategies for doing research and how to collaborate in the research environment with their colleagues as well as with the society at large. Scientific information strategies should support and therefore facilitate these researchers' strategies.

In this vein, a proper starting point is to first discuss research using the concept of the business model as guidance to analyse the research environment, competition in research and drivers in research for making research results public and for acquiring these results by other researchers. This allows discussing criteria for business models in the information market and developing scenarios for scientific information and their consequences for all stakeholders, researchers, publishers, librarians alike. It allows speculating on the consequences for the business model of research and HE institutions as e-science opens up new possibilities for collaborations in projects across such institutions. In particular, it will create new challenges for the smaller and medium institutions to participate in such collaborations.

1.2 From value chains to business models

Changes in the value chain are triggered by engines of change.[6] *For this market these engines for change are the potential that ICT offers to empower the author and reader and the recent developments in research and HE, also to a large extent but not exclusively enabled by the potential offered by ICT. ICT provides a huge potential to empower the author and the reader and allows a change from a use-oriented system towards a more availability-oriented system at the same time allowing a new balance between centralised systems and distributed or federated systems. ICT raises for the stakeholders the strategic choice between empowerment of the user, or alternatively applying a hostage strategy directed at the user in particular.*[7]

With respect to some broader developments in research it may suffice to mention that research has generally become more subject to market conditions, even when carried out in the environment of a research institution. Market conditions mean that intellectual capital and scarcity of resources, both financial and human, play a more and more important role. As a result, research information is being intensively used for planning and evaluating of entire research programmes em-

6 Roosendaal H.E. (2004) Driving Change in the Research and HE Information Market. *Learned Publishing,* vol. 17, no. 1., p. (…)

7 Freeman E., Liedtka J. (1997) Stakeholder Capitalism and the Value Chain. *European Management Journal*, vol. 15, no. 3, p. 286-296.

phasising the formal publication side system rather than the communication side. This means to say that the balance between real communication between researchers as opposed to formal publication of research information is even more changing to formal publication.

In education, the introduction of the bachelor/master structure at the European universities will spur the development of web-based and blended learning when students are becoming more mobile and will hop from one university to another. This mobility is expected to show up in particular for master students and will lead to the introduction of international masters. Wider applications of distance learning and life-long learning will spur these developments.

For our discussion it is interesting to note that the information requirements - in terms of publishing and archiving - for research and for educational materials are very similar indeed. For educational information the volume required for each HE institution is at least an order of magnitude larger than the research information it requires. This makes it attractive from an institutional point of view to have research information financially piggy-backing on educational information. HE institutions have to develop their information infrastructure for the production and registration, i.e. publishing and archiving of educational material anyway and can use that infrastructure for the production and registration, i.e. publishing and archiving of research information as well. In both cases this includes the production and registration, i.e. publishing and archiving of grey information.

Nevertheless, rather than focusing on engines for change and the value chain a more comprehensive argument based on the business model for the scientific information should be used.[8]

Any business model should serve the following conditions:

- It should create value in its environment[9] in the process at hand, i.e. the production and sharing of knowledge.
- It should create a sustainable process.
- It should create value for commerce.

A business model is thus viewed as the organisation of property and of the exchange of property, the property being the knowledge produced by the researcher and in particular the intellectual property of this researcher, as well as the added value of all other stakeholders.

Following Chesbrough & Rosenbloom[10], a business model

- articulates the value proposition;
- clearly defines the market segment;
- reflects the strategic position of the researcher;

8 See Roosendaal et al. (2009), op.cit.

9 Kurek K., Geurts P.A.Th.M., Roosendaal H.E. (2006). The split between availability and selection. Business models for scientific information, and the scientific process? *Information Services & Use*, vol. 26, no. 4, p. 217- 282.

10 Chesbrough H., Rosenbloom R.S. (2002) The role of the business model in capturing value from innovation: evidence from Xerox Corporation's technology spin-offs companies. *Industrial and Corporate Change* vol. 11, no. 3, p. 529-555.

- identifies the value chain in the market of scientific information;
- reflects researchers' competitive strategy;
- identifies revenues and costs structure and profit potential.

A major boundary condition is that business models for the scientific information market should be commensurate with the research environment in order to serve research. Under this condition, the main parameters of a business model for the scientific information market have been shown to be the availability of scientific information and the power of selection of the researcher.

Relevant for the discussion on engines of change is the notion that making research results public is an important tool for researchers to position themselves in their environment, the research environment and the wider societal environment. It is for this reason that a brief discussion of the concept of strategic positioning is given here, as this positioning is relevant for establishing a strategic relation resulting in the production of knowledge to be made public.[11]

Researchers establish such a strategic relation with their environment with the goal to create added value. Partners decide to collaborate because in a situation in which they would not have access to resources of other researchers they would not be able to create added value and to achieve their goals. Establishing this strategic relation is essentially a process of acquisition of resources and negotiation between these two partners on sharing heterogeneously distributed strategic resources and on governing the directions of research. Researchers decide to give up governing research to a certain degree and accept sharing resources to a certain degree.

From the literature, a number of modes of strategic positioning is known. In "mode1", researchers set research directions driven by scientific curiosity. Results of research are not necessarily meant to be of societal relevance. Therefore, researchers can restrict the communication and collaboration to their research environment. In this case, researchers do not need to influence this environment. This type of positioning is well-known as 'ivory tower' or 'free research'.[12]

In the so-called "mode2", the societal environment directs researchers. It influences research directions and ipso facto influences the scientific products they deliver. This means that researchers match their own research problems to existing research programmes based on the demand of the societal environment. They are "context-sensitive",[13] listen to the environment and fulfil societal needs.

---

11 A more extended discussion can be found in Kurek K., Geurts P.A.Th.M., Roosendaal H.E. (2007). The research entrepreneur: strategic positioning of the researcher in his societal environment. *Science and Public Policy*, vol. 34, no. 7, p. 501-513.

12 Ziman J. (1994) *Prometheus bound. Science in a dynamic steady state*. University Press, Cambridge.

13 Novotny H., Scott P., Gibbons M., (2003) Introduction: 'Mode2' revisited: The New Production of Knowledge. *Minerva* vol. 41, p. 179-194. See also Gibbons M., Limoges C., Novotny H., Schwartzman S., Scott P., Trow M., (1994) *The new production of knowledge. The dynamics of science and research in contemporary societies.* SAGE Publications, Stockholm.

The "mode3" position introduced by Kurek et al. (2007)[14] means that researchers share resources with the environment like "mode2" researchers. But contrary to "mode2" researchers, "mode3" researchers or "research entrepreneurs" have the opportunity to be autonomous in determining the directions of research. They retain their own responsibilities for directing a project. Research entrepreneurs, like business entrepreneurs, influence the societal environment by creating demand for their scientific products. "Mode3" is seen to be compatible with e-science in the sense that e-science facilitates "mode3".

One aspect of scientific information, such as information for any business organisation, is to create competitive advantage for the research enterprise. Competitive advantage based on scientific information enhances the influence of researchers not only in their research environment but also leads to a better strategic position in the societal environment. For this very reason, it is important to deal in a succinct way with this aspect of competition as an engine of change, in particular how it relates to making research results public and in acquiring scientific information. This is particularly relevant when competition is changing due to a change in researchers' modes as facilitated by e-science.[15]

1.3 Functions in scientific information

As stated above, the driving force for the market of scientific information is that "authors want to publish more" and have their product widely available, while "readers want to read less", but want to be informed of all that is relevant for their research at hand. Readers want this information available just in time. They want to be guaranteed that they can and will be informed of all that is relevant to them.

This market thus consists of researchers as producers of knowledge (authors) and as users (such as readers) of knowledge, the overall goal of researchers being to produce knowledge. Moreover, in the process of production of knowledge they acquire and make use of scientific information produced by others. Therefore, discussing the market means discussing the combination of the production of knowledge and the acquisition of scientific information.

Next to researchers and other stakeholders such as libraries, digital networks, publishers, and agents etc. the market consists of the product of scientific information, as the objective of researchers is to share scientific information. As we know, researchers are not only producers but also heavy users of scientific information produced by others. The condition here is that scientific information must have been made public.

---

14 Kurek K. et al. (2007) op. cit.

15 A more extended discussion can be found in Kasia Zalewska-Kurek, Peter A.T.M. Geurts, Hans E. Roosendaal,(2008). 'The role of business models for scientific publishing in the research environment', chapter4 in Kasia Zalewska-Kurek *Strategies in the production and dissemination of knowledge.* PhD dissertation. University of Twente.

Forces that can be observed in this market are therefore related to researchers and scientific information itself. The driving force for researchers in producing scientific knowledge is recognition. Important motives to publish research results have been seen to be recognition and visibility.

Recognition leads to reputation and researchers report produced knowledge as an instrument in the acquisition of resources. The goal is to be recognised and competition is the organisation of actions and efforts of researchers to attain this goal of recognition. Recognition and competition are attributes of the researchers and availability and selection are attributes of the product. Researchers in the market of scientific information require knowledge that can be easily acquired. It has to be available and easy to select. Only in this way researchers gain a competitive advantage in competing with other researchers. The forces are complementary and should be properly balanced with regard to the researchers and their positioning in the environment.

Following these arguments, one can deduce that the driving forces in the scientific information market are recognition, competition, availability and selection. The main functions of scientific information are then registration, awareness, certification and archiving (Figure 1.1).

These functions are defined as strategic functions from a science point of view.[16] *The external functions registration and archiving are seen to be outsourced out of science to the publisher and the library respectively.*

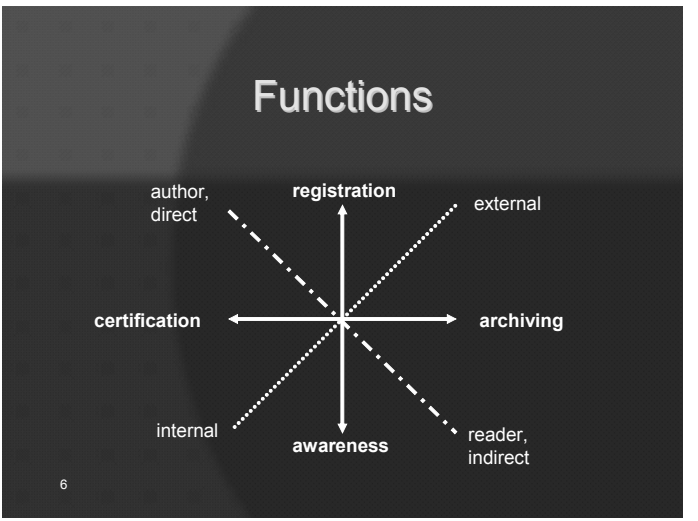

Figure 1.1 Strategic functions of scientific information

The four functions in scientific information need always be performed independently of the technological environment, albeit that the balance between the func-

16 See Roosendaal et al. (2001), op.cit.

tions may well change under changing technological conditions. We will use the four function scheme for scientific information as analytical tool in our analysis of changes in the value chain arising from the engines of change as discussed above. They provide amongst others a powerful check on the comprehensiveness of these scenarios and the consequences for the stakeholders.

Sharing information is the main value proposition that any business model should account for: it should allow researchers making research results public and acquiring scientific information. As the intellectual property is the main property in scientific information, any business model can only serve researchers in producing knowledge if it serves the author in claiming intellectual property next to serving the reader in acquiring scientific information. This can only be achieved by guaranteeing adequate availability of scientific information. In addition, the ability to acquire scientific information depends necessarily on the availability of such information next to the ability of selecting this information by the researchers. This means that the information should in principle be universally accessible.

In the above, we have implicitly defined the market segment as the research environment worldwide. In the narrower sense, this implies that the reader will want to acquire information and to use this information to do further research to produce future research results. This seems the main use of scientific information, but scientific information is also used in areas of application outside the original research area. Such areas of application can be other research areas, interdisciplinary areas or even application outside research, e.g. in societal applications, such as in industry, services or the public at large. This means that the market segment is clearly broader than the research environment. Nonetheless, the main objective remains to share information and it is therefore the receiving end that determines how to make use of this information for their goals and purposes. A main observation to add then is that the value proposition is therefore in principle determined by the demand side.

1.4 Value chain options

As stated above, ICT in particular allows a variety of value chains. The value chain is defined being linear in terms of steps of added value and is not a process chain. The corresponding process chain is in essence a rather complex network of process steps.

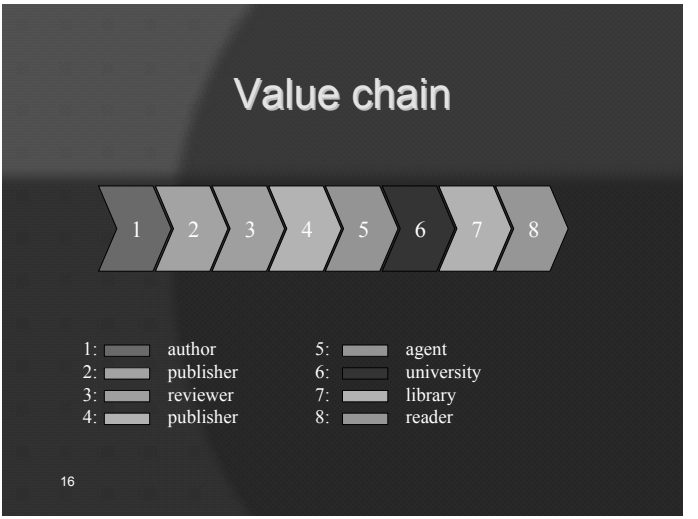

Figure 1.2 Traditional value chain

Figure 1.2 shows the traditional value chain, as we know it from the paper-based environment. In this figure we show the value chain with the stakeholders responsible for the added value per link. Thus the author creates the work, sends it to the editor, the publisher will produce the work and send it to the university. Administrative assistance is mostly given by an agent. Finally the paper arrives at the reader.

1.4.1 Alternative options

In Figure 1.3 we show a shortened value chain of author and reader only, i.e. full empowerment for the author and the reader. This means no quality filter or branding. This value chain can well work for information that the reader is very familiar with, but takes an extraordinary effort on the part of the reader with information less familiar thereby violating the statement: 'Authors want to publish more, readers want to read less.' This value chain is totally availability-based meaning that the author or the institution does not only have to bear the financial risk but as there is no refereeing there is also for the author the full risk as scientific entrepreneur.

Figure 1.3 Value chain with full empowerment for author and reader

Another possible value chain is the one in Figure 1.4 where publishers are delivering information directly to the reader. Weak point in this value chain is the responsibility for the archive that in this case should rest with the publisher, not a very realistic proposition. This value chain is totally reading use based and costs will have to be picked up by the reader.

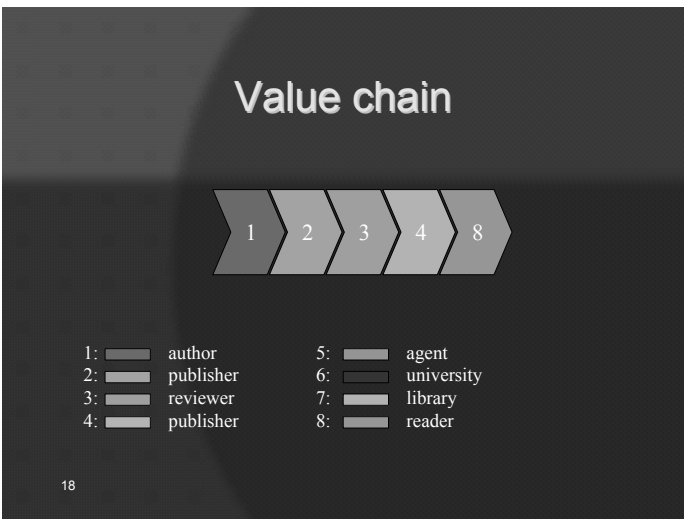

Figure 1.4 Value chain without universities

Alternatively, we could swap the publishers for the HE institutions taking over the publishing function (Figure 1.5). In the case of research information the weak point then is the certification of the material. This cannot be managed by the home institution of the author. A way out could well be the creation of alliances of institutions, leading finally to the establishment of new publishers. However, for learning material this value chain is highly feasible as in this case the 'buying' institution can exercise the certification power.

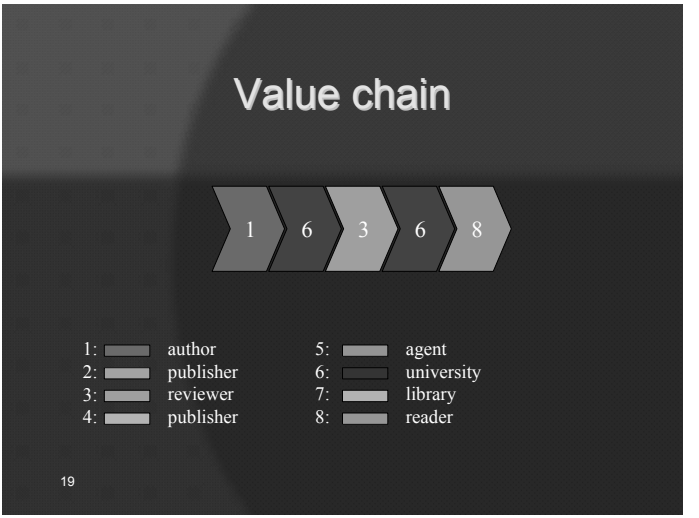

Figure 1.5 Value chain without publishers

In the last figure we see a value chain that looks rather similar to the traditional value chain, but with totally new roles for the stakeholders. The institutions are now responsible that the work (author) can be sustainably archived ('perpetual' archiving) and is properly disseminated to the reader. The institutions are in this chain responsible for the registration and the archiving functions (Figure 1.6). The publisher is responsible for the distribution and branding and in providing logistical assistance for the editor in the certification process.

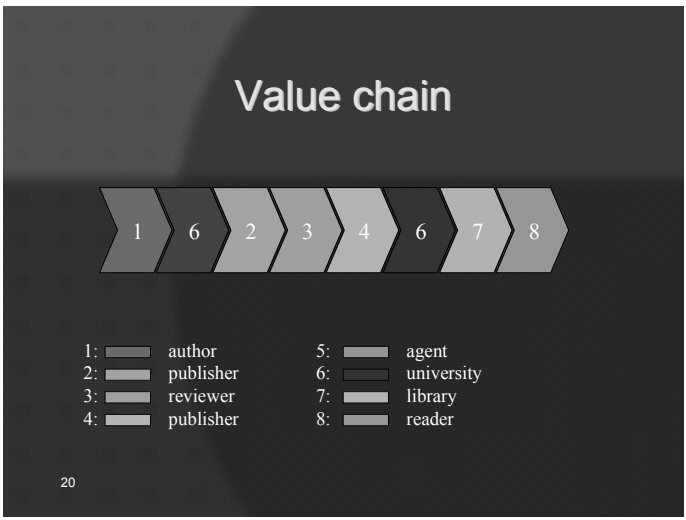

Figure 1.6 Value chain with new roles for institutions and publishers

This chain has a number of consequences:

1. *The fixed first copy costs have not to be born twice, i.e. by the institution and the publisher, but will be born by the institution only.*
2. *The author can transfer copyright, i. e. the overall exploitation rights, to the institution where the work was performed. The institution can then transfer specifically designated rights to the publisher.*

In this last value chain the costs for production and dissemination will be born by the institution. The different options of the value chain allow different options for scenarios in the research and HE information market. The different options represent also differences in the balance between availability on the one-hand side and reading use on the other hand. This is relevant for the different business models emerging from these options.

1.4.2 Business models

As we have seen before, the value chain of all stakeholders involved in the entire process should be a part of the business model. A major consideration then is that if serving researchers is the main value proposition, any business model should account for the conditions determining how researchers are conducting research. This means that this model should account for the different modes of strategic positioning in which different types of scientific information is being required, acquired and produced. This then results in requirements for the value chain.

The business model should account for competition within the research environment that, as argued, affects the researchers' choices, requirements, and the necessary conditions for scientific information. Part of the competition is in claiming intellectual property which evidently creates a competitive advantage for the owner of that property. But there is also a competitive element in the acquisition of information.

Full availability of information can be argued to be of particular relevance to smaller research institutes as they are necessarily more limited in their networks and generate less knowledge than larger institutes. Medium and smaller research institutes may therefore be more vulnerable for limited availability of information as this may hamper them in producing new knowledge. Effective acquisition of scientific information also depends on the power of selection by researchers. This power of selection, possibly enhanced by various services, gives researchers additional competitive advantage in terms of improved access to relevant and up to date information acquired at the right time.

As we have noted above, a business model should provide a proper balance between availability of scientific information and selection of this information by researchers. A proper balance influences the researchers' ability to acquire and select relevant scientific information and therefore, impacts on their competitive advantage. Grey literature can provide an important service in this respect.

The revenues and costs structure and profit potential in the business model is shown to be dependent on the organisation of the two main dimensions that we have noted before: availability and selection, or rather the balance between these two dimensions.

Another condition is that the business model should be sustainable, where sustainability is defined as the characteristic of a process, system or state that can be maintained at a commensurate level, and in 'perpetuity'. This boundary condition is seen to be particularly relevant in scientific information in its service to the production of knowledge with its strong demand for legacy. The boundary condition of sustainability means that scientific information should be available and accessible in perpetuity at the same time requiring a revenue, costs and profit structure that can ensure this demand. It may be noted that a subsidised and therefore political system, would not possibly only render the scientific information system very vulnerable, but could also endanger independent certification of the research results, in this way endangering the research process itself. Sustainability and its consequences are issues that also grey literature should account for.

Another issue that grey literature should deal with is peer review. Peer review certifies the researchers' contribution to scientific knowledge and 'brands' it. In the process of peer review the research environment decides if the claim to the property by the author can be made, if the claim is of commensurate scientific value. Being essential for claiming the property, peer review is therefore core to any business model for scientific information.

Any business model is based on a combination of the two parameters of availability and selection. Neither the subscription model nor the open access model does entirely fulfil the necessary conditions for general availability and power of

selection at the discretion of researchers. Each of these models focuses too much on one parameter.

The business model for grey literature represents a family of variations of the optional business model, a characteristic being that the registration and archiving functions are combined in the author's institution.[17]

Another conclusion is that the dominant business model, the subscription model, is heavily supply oriented while providing bounded or limited availability and in doing so is in principle a publisher centred model while at the same time focusing on the author as the primary stakeholder for consideration. The open access model in all its variations as coming up in the market is in essence also a supply oriented model. It is furthermore like the subscription model primarily a publisher centred model, in particular in its forms of open access mandates for publishing on the institution's repository followed by subsequent publication in a journal.

This means that both known business models, the dominant subscription model and the open access model both in their different manifestations in the market are essentially supply oriented and publisher centred, whereas convergence of the scientific information market towards e-science can only result in a business model that should be demand oriented and above all research centred.

Demand oriented means that the business model should fulfil the demand of authors for full availability and the demand of readers to decide on their own needs for selection depending on the information they want to acquire. Research centred means that the business model should allow for the different strategies researchers want to develop in strategic positioning themselves in the relevant environments and for competing in these environments. Any business model, grey or not, should comply with the prime demand of research of sharing scientific information for the benefit of research, i.e. sharing information in a very dynamic environment demanding that information must be made public and can be fully acquired.

1.4.3 High-level strategy

Creating a network of repositories of information relating to research and education requires a basic conception of a high level strategy shared between the different stakeholders having different business philosophies. Such a strategy can only be successful if it fulfils in the best possible way the major interests of the stakeholders. This requirement means that such a strategy can only have one focus: the user as the primary beneficiary of the network. This is the only possible strategy leading to value creation, the alternative being value capture by one of the stakeholders and taking the other stakeholders, in particular the users, as hostages. The user is the learners, teachers, researchers and students in knowledge institutions and organisations, in their capacities as author and/or reader. This means

17 See Roosendaal et al. (2009), op.cit.

that a comprehensive approach to user behaviour and to the consequences of such behaviour for the value chain of information is indispensable.

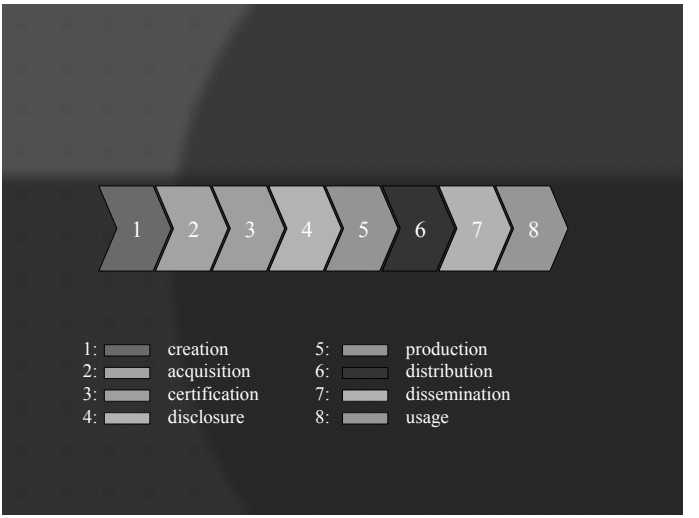

Figure 1.7 Strategic tasks of the network as represented in the value chain

The institutional stakeholders in the research and HE information market and beyond, will as enablers be the secondary beneficiaries. As stated before, the foremost goal for every stakeholder is to develop an individually tailored strategy to comply with the high level strategy in this way positioning this stakeholder at the forefront of developments in on-line information management. Only then the stakeholder will be able to make an invaluable contribution to a network for worldwide information provision in research and education. A key aim of this strategy is making universities and other knowledge institutions, scientific publishers, non-commercial or commercial, professional by helping to make use of this network and ensuring that the architecture will best serve all stakeholders' needs. The network should be able to support the user in the strategic tasks as embedded in the value chain in Figure 1.7.

As a consequence of such a high level strategy the corresponding technology strategy should focus on developing an architecture for federating existing and future repositories and libraries for the familiar strategic reasons for making use of an architecture:

- *to reduce complexity;*
- *to allow a proper balance between central and decentral aspects of the development;*
- *to be able to manage change properly;*
- *to facilitate experimentation and competition;*
- *to ensure that many different systems can develop together gracefully.*

A main goal for this architecture is the development of a shared architecture for e-documents, e-learning and e-science and this requires integration and resources syndication. A foremost strategic goal is that the authentic copy of a work of whatever type, should remain located at the home repository, being the repository of the affiliation of the creator of this work. This would constitute an important step towards empowerment of the user.

The relation between the research environment and the information environment, i.e. research and HE institutions, repositories, publishers and other intermediaries, requires a sort of virtual organisation comprising of these two environments as to ensure steady progress in the development towards e-science. In fact, it calls for a sort of organisation like we know well from the development of the World Wide Web: the WWW consortium. In this way, a worldwide scientific information network as described in the vision could be realised with a dispersed spectrum of stakeholders ensuring a diversified and differentiated network that is optimally integrated in research and teaching.

1.5 Concluding remarks

In further analysing the consequences for stakeholders including researchers and research and HE institutions in a way consistent with the above discussion, we again have to look at the production of scientific information as an alliance or as a sort of integration of the main stakeholders with the research environment. This seems valid as it is evident that research centred and demand oriented business models require some degree of integration between the stakeholders. Here, a grey business model could possibly be advantageous as this business model is per definition more integrated.

We conclude our chapter with some summarizing remarks.[18]

- Researchers demand a research centred and demand oriented family of business models for scientific information as only such models ensure that scientific information serves the production of knowledge, results from the side of the. These business models ensure further integration of scientific information into the research and teaching enterprise in its development towards e-science.
- As for research and HE institutions it is evident that high value information provision is a strategic core activity of every institution and becomes even more relevant in the development towards e-science. Institution management has to be aware of this responsibility for the provision of adequate information services.
- The research and HE institutions are the natural candidates to initiate the development of new business models and structures. This is foremost an organisational and not a technical challenge. A major organisational chal-

18 See Roosendaal et al. (2009), op.cit.

lenge will be to absorb the library consequently into the research organisation. The goal of this absorption is to change the relation between the institution's primary processes and the information provision for these processes. It has been seen necessary that this information provision will have to integrate more closely with the primary processes to deliver the services they need.

- The developments in the market of scientific information, in particular the convergence towards e-science provide great opportunities for professional, commercial or non-commercial service providers. To grasp these opportunities it is important that these service providers will develop a more integrated relation with the research environment.
- Other service providers will have opportunities to assume tasks to support the functioning of the overall network. Tasks can be in the areas of technical and administrative support. There is a special task in controlling the logistics of the network.

Any business model should comply with the prime demand of research of sharing scientific information for the benefit of research, i.e. sharing information in a very dynamic environment demanding that information must be made public and can be fully acquired. Such a business model leads to a network comprising the research environment as the pivotal stakeholders together with the other stakeholders. Such a network requires careful strategic positioning of these other stakeholders with respect to the research environment. As stated above, a grey business model is a family of variations of the optional business model.

The technical and organisational development bears important consequences for the strategic development and use of grey information. Rather than seeing grey literature as type of product or a set of types of products it may well be tempting to consider grey literature as a specific type of value chain(business model) or a set of specific types of value chains (business models) in the entire family of value chains (business models) possible in information related to science.

Indeed, in grey literature the registration and archiving function have always been combined at the author's institution, being this an individual author or the institution itself. Grey information that is published on the institutional repository will then enjoy wide availability as opposed to limited distribution as used to be the case and this will make grey information straight away the most abundantly available scientific information.

The challenge for grey literature is then to find ways to integrate fully into the further and continuing convergence towards e-science.

Acknowledgement

The author should like to thank Dominic J. Farace and Joachim Schöpfel for inviting him to write this article. The author should like to thank Joachim Schöpfel for his encouragements and invaluable help to write this article in this particular form.

Chapter 2

How to assure the Quality of Grey Literature: the Case of Evaluation Reports

Markus Weber, Federal Office of Public Health, Switzerland

2.1 Introduction: Grey Literature needs quality control

The production of grey literature has grown considerably compared with the more traditional type of academic literature. This is due to several developments in modern society and many of them are covered by the two key words «knowledge society» and «internet». One consequence of this is that today readers have more difficulty in judging the quality and relevance of what they read. This is especially true for information professionals in all sectors of government, academics, business and industry and stays a major challenge for developments in our information society.

At first this seems to be mainly a problem on the demand side of grey literature (the reader), but is of course also a problem perceived and being tackled by the supply side: How can one be assured of the quality of grey literature in a similar way to the quality hallmark of "white literature" i.e. peer reviewing in scientific journals?

This chapter is about a quality assurance system[1] for a specific category of grey literature, evaluation.[2] It was developed by the Competence Centre for Evaluation (CCE)[3] of the Swiss Federal Office of Public Health (FOPH) as a six

1 «Quality assurance», «quality control» and «quality management» are used as synonyms in this text.

2 «Evaluation is the process of determining the value (contribution to societal well-being), quality, and/or justification of the object in question. Its judgment is based on the use of (mostly) social science research methods and procedures for the systematic collection and analysis of data, not necessarily routinely available, regarding various aspects of a public measure. The judgment criteria most commonly applied include RELEVANCE, EFFECTIVENESS, and EFFICIENCY, and occasionally, SUSTAINABILITY.» (FOPH Swiss Federal Office of Public Health 2005)

3 The CCE is responsible for commissioning and managing the FOPH's evaluations of public health measures - mostly of health promotion and prevention programmes and projects[3]. It is an internal service that has to assure the studies' scientific quality, ethical conduct and trustworthiness, on the one side, and, on the other, their usefulness. Most studies are mandated to external university research institutes or private evaluation consultancies.

step system for assuring the quality of the commissioning and management process as well as the evaluation products, especially the written reports. The "effects" of such work have been well recognised; «Managed by the CCE» is increasingly perceived in Switzerland as a quality label for evaluation studies. For example, The Swiss FOPH and its CCE are mentioned in several international and national studies as a good and successful example of how to handle evaluation in public administration (e.g. Fornerod 2001; Jacob and Varone 2002; Widmer et al. 2001).

The CCE's quality assurance system is described in detail in section 2.2. Whilst much of its experience is concerned with public health evaluations, the system itself could probably also be applied to other areas concerned with producing knowledge for grey literature. The implications of such a transfer are discussed in 2.3, and some conclusions are presented in section 2.4.

2.2 An example of a quality assurance system for commissioned evaluation studies

2.2.1 Overview and objectives of the system

By introducing and using a quality assurance system the CCE aims at achieving two main objectives; firstly, that the studies are conducted according to sound evaluation standards, including the scientific quality of the applied methods and methodology and secondly that the products are useful and practicable, i.e. the studies need to address questions and draw conclusions that are relevant to the needs of a wide and varied audience, and come up with a set of recommendations that can be implemented.

2.2.2 General description

In most cases, the final product of an evaluation takes the form of a written report; quality control is most often therefore focused on this end product. However, evaluation is a process as well as a product and thus there are many steps along the way that need to be controlled for quality. It would not be very sensible to just come in at the end of a study and judge the quality of a report; rather it has to be steered from the beginning. The CCE has standardised processes, guidelines, models and checklists that are used to guide the process from A to Z, i.e. from the first request for a study to actual commissioning, accompanying the study throughout, assessing the report (meta-evaluation) and discussing and supporting a work plan for the utilisation/implementation of the study results.

Figure 1 shows the 6 main steps of the evaluation process from a commissioner's point of view. The many sub tasks that have to be considered within each

step are described on the following pages. Many of these are supported by CCE checklists, models, etc.

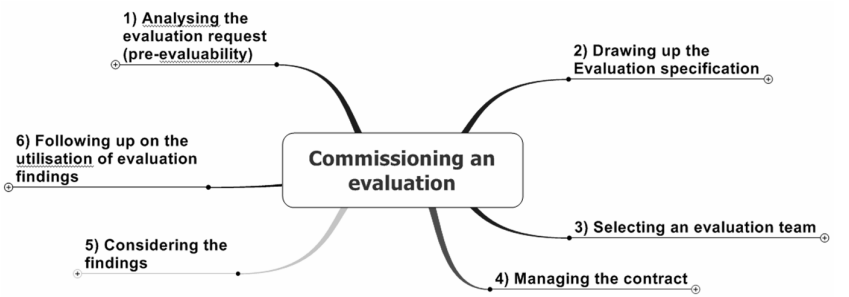

Figure 1: Commissioning an evaluation in 6 steps

The process starts with the CCE's analysis of an evaluation request that it receives from the specialist internal service needing the evaluation (e.g. the HIV/AIDS prevention unit). If the request is considered justified and necessary the CCE then develops the evaluation specification (step 2). After calling for offers, the CCE, together with its internal partner, selects an external evaluation team (step 3). The study is then commissioned and the CCE regularly meets with the external team, reviews the tools developed for data collection and analysis and manages the contract (step 4). At the end (and sometimes midterm) the study's findings are received, considered and a plan of action is drawn up to put them to effective use (step 5). The final part of the process includes following up on the "action plan" about one year later to see what was done and what was achieved

This system has been successfully used for several years (FOPH Swiss Federal Office of Public Health 1997; Läubli Loud 2004) and, more recently, adjusted to take into account the Swiss Evaluation Society's (SEVAL) quality Evaluation Standards (Widmer et al. 2000)[4]. As SEVAL's standards underpin several aspects of the CCE's quality control system, they and their role are explained in the next section (2.2.3), before going on to present the details of the CCE's 6 steps (2.2.4).

2.2.3 The SEVAL Standards

The SEVAL developed its set of good practice standards for guiding the conduct of evaluations in Switzerland. They refer to the processes involved in seeking and collecting data for making judgements and producing the written report. They describe what an ideal evaluation should be like in an ideal situation. They also promote the need for self-reflection and professional discussion between commissioners, evaluators and any other stakeholders so as to build a common ground for

4 Approved in 2001 by the Swiss Evaluation Society SEVAL, www.seval.ch [site visited 04.08.2009]

the execution of a study. As such it is hoped that the risk of a study being instrumentalized or manipulated is reduced.[5]

The 27 standards are grouped into four quality dimensions: Utility, Feasibility, Propriety and Accuracy (cf. Figure 2). The objective of each dimension is as follows (Widmer 2005):

- The **8 Utility standards (U)** guarantee that an evaluation is oriented to the information needs of the intended users of the evaluation
- The **3 Feasibility standards (F)** ensure that an evaluation is conducted in a realistic, well-considered, diplomatic and cost-conscious manner
- The **6 Propriety standards (P)** ensure that an evaluation is carried out in a legal and ethical manner and that the welfare of the stakeholders is given due attention
- The **10 Accuracy standards (A)** ensure that an evaluation produces and disseminates valid and usable information

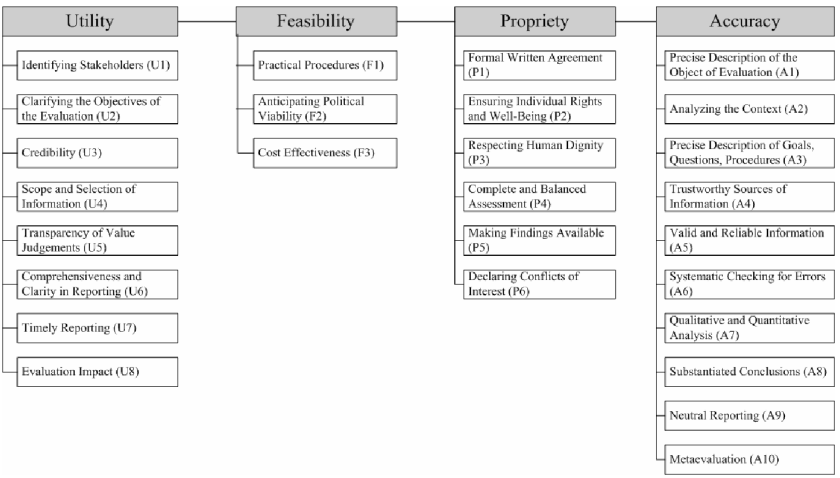

Figure 2: The 27 standards ordered by the 4 quality dimensions (Widmer 2005).

The SEVAL Standards define the expectations of an evaluation but do not specify either the methodology or the methods to be used. Overall, they share the same concerns and objectives as those defined by the CCE: sound scientific quality and ethical conduct (especially through the accuracy and propriety standards) and production of practical knowledge (utility and feasibility standards). The standards are categorised according to the quality dimensions. But they are not all equally relevant to every evaluation (e.g. subject to which methodology is applied) and certainly not to every phase of an evaluation (from initial planning to utilisation).

5 This common ground is usually referred to as an «evaluation culture». A more detailed description on the background of an «evaluation culture» and on the current status quo of its development in Switzerland can be found in Läubli Loud (2004)

Those who use the SEVAL Standards need to relate them to their specific evaluation methods, needs and situation. The head of the CCE team was involved in the development of the SEVAL Evaluation Standards (Widmer et al., 2000) and has systematically advocated their use in the commissioning process ever since.[6]

2.2.4 Step by step through the CCE's system

The step by step description of the CCE's quality assurance system provides a good overview of how each part of the system works. Even though all of the 6 steps consist of several sub-steps (cf. 3 to fig. 8), only those of key importance to quality assurance are described in further detail.

Step 1: Analysing the evaluation request (pre-evaluability)

At this point, a CCE staff member is asked to study the request and determine its main aim or purpose, the key evaluation "needers and users", and whether it is worthwhile and feasible, e.g. can the information be obtained in some other way such as through performance review, or audit and/or can the expected information be delivered "in time" enough to be useful i.e. to help decision making? Background knowledge has to be gathered and processed to help clarify the main purpose of the requested study and the intentions and expectations of different purposes of different stakeholders.

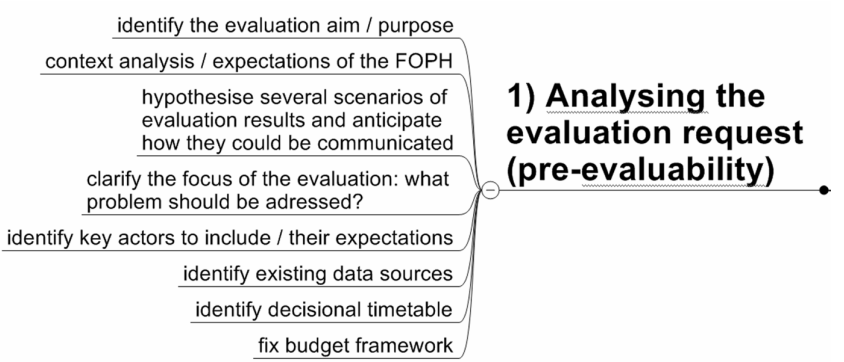

Figure 3: Step 1, Analysing the evaluation request (pre-evaluability)

Step 2: Drawing up the Evaluation specification

This is a very important step since it makes a considerable contribution to determining the final quality of the study. It sets out the key questions that need to be addressed, the scope and focus of the study, who needs what information and how

6 A "slimmed down" version of the SEVAL standards was later produced specifically for the use of commissioners within the Swiss federal administration (Widmer 2005).

and by whom it is intended to use the findings, and most importantly, the time frame for receiving the results.

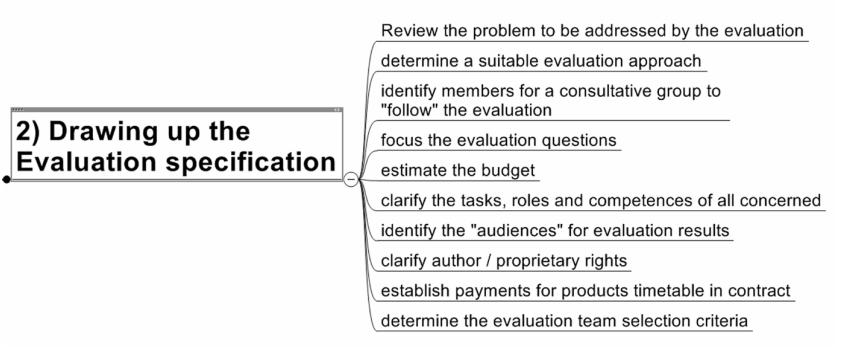

Figure 4: Step 2, Drawing up the evaluation specification

Step 3: Selecting an evaluation team

Once the evaluation specification is finalised the study is put to tender and the CCE, together with the internal specialist service needing the evaluation, examine the offers and select a suitable evaluation team.

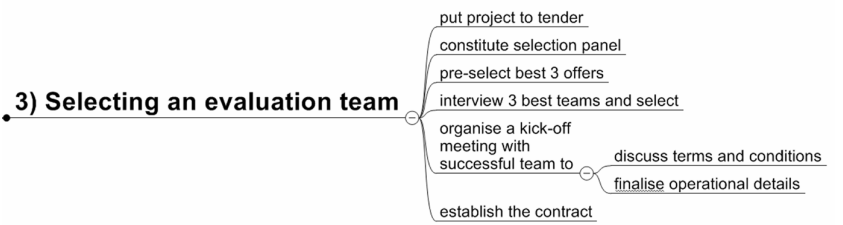

Figure 5: Steps 3, Selecting an evaluation team

Step 4: Managing the contract

In this step the CCE staff member responsible for managing the project keeps in regular contact with the external evaluators for monitoring progress, reviewing the tools, helping with gaining access to data and organising regular feedback sessions. Often intermediate results can already be very useful to the internal commissioners of the study (end user). Such regular contacts are therefore essential for identifying and bringing forward useful information “along the way”.

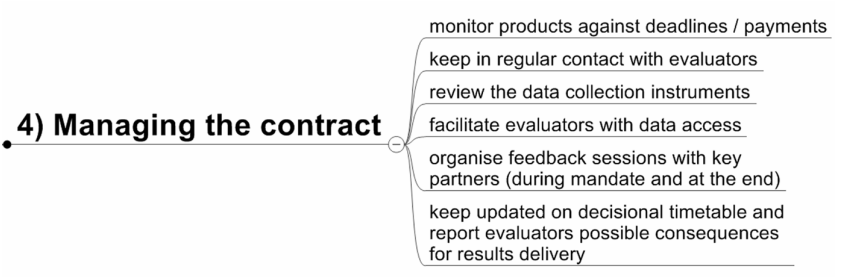

Figure 6: Step 4, Managing the contract

Step 5: Considering the findings

In step 5 the evaluation's findings, conclusions and recommendations are presented in both written and oral form to the commissioners and other stakeholders. However, the quality of the report is first checked by the CCE and this "meta-evaluation"[7] phase is the most important for assuring the overall quality of the work. As the end users have to take responsibility for interpreting the findings (what do these mean to their work?) and developing a dissemination and action plan (what has to be done consequently?), they depend on the CCE's quality control of the overall evaluation

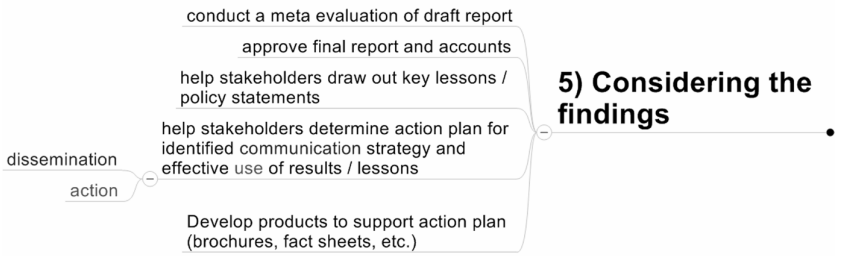

Figure 7: Step 5, Considering the findings

A set of checklists supports the CCE in its examination and evaluation of the final report. Step 5 is therefore the second most important and time consuming task of the overall system.

Step 6: Following up on the utilisation of evaluation findings

This step is less relevant for the CCE's control of an individual study, but more for the control and accountability of its overall products and services. It is of paramount importance for legitimizing evaluation in an institution such as the

7 A meta-evaluation is defined by the FOPH as the scientific and ethical quality control of an evaluation (FOPH Swiss Federal Office of Public Health 2005).

FOPH. Evidence on the usefulness of evaluation per se and of CCE's work in general has to be demonstrated – the CCE therefore seeks out evidence on how the knowledge and lessons highlighted through the evaluation studies have helped in making further decisions, fundamental changes and/or slight modifications to the public health strategies and measures studied. Towards this end, it compares and contrasts the intended "action plan" with what was finally implemented and why.

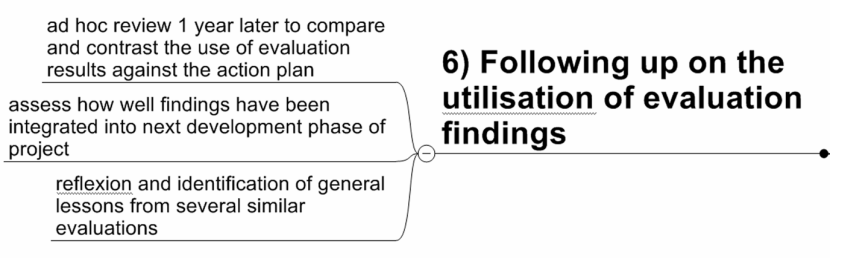

Figure 8: Step 6, Following up on the utilisation of evaluation findings

2.2.5 Short discussion of the quality assurance system

The 6 step system's procedures and tools help the CCE to achieve its two main objectives (assuring its evaluation studies are of sound scientific quality and produce useful and usable knowledge which can be put to practical use). Scientific quality and professional ethical conduct is assessed through a strict review of the final product (meta evaluation of the final evaluation report) - the last of the quality assurance procedures. Steps 1, 2 and 4 are the most staff resource-consuming; however, given the guidelines and checklists produced to support each step along the way, the 6 steps of the system can now be accomplished in much less time than was possible before.

2.3 What for other study types and constellations?

In this last part of the chapter we have described the system used to help the CCE assure its partners of quality evaluations. The procedures are based on some general principles of quality assurance and therefore should be readily transferable to other "grey" literature areas. Below, we suggest some possible ways of transferring "good practice" principles to other areas albeit adapted to the needs of other contexts.[8]

8 Perrin (2006) argues that in most cases speaking of «best practice» should be avoided. Before transferring «some successful way of doing something somewhere» to your own

For evaluators: adapt slightly

For evaluators themselves, much of the CCE's system for overseeing the commissioning process cannot be applied directly. However, the SEVAL evaluation standards address evaluators and commissioners alike as they set out standards to be addressed both during the process and when producing written reports. In step 5, the evaluation team should therefore conduct a meta-evaluation of their own work.

Other kinds of studies than evaluations: Use other, relevant standards for quality assessment or …

In many other areas standards of good practice exist and can act as a starting point for developing a quality assurance system e.g. in clinical research one could take aspects from the CCE's 6-step-model (or any other defined process) and combine them with the quality standards of "good laboratory practice GLP" of the OECD[9].

… use GLISC guidelines as an instrument

A very useful tool is the «Guidelines for the production of scientific and technical reports: how to write and distribute grey literature», also called «Nancy style» (GLISC 2007). These guidelines are mainly focused on writing (and distributing) accurate, clear and easily accessible scientific reports in different fields, but they also «include ethical principles related to the process of evaluating, improving, and making available reports, and the relationship between GL producers and authors» (GLISC 2007, p. 1). «The Guidelines state the ethical principles in the conduct and reporting of research and provide recommendations relating to specific elements of editing and writing» (GLISC 2007, p. 2).

Minimal procedure: Clarify everything along process

The main element of the CCE's system is of course the same as for any good research: clear, transparent and well documented procedures for guiding the procedures and conduct of the work.

2.4 Conclusions

The quality assurance system described above was specifically developed by the CCE to oversee the evaluation process and product. The system is based on the fundamental premise that the "end" is the result of the "means" used to get there. Thus the quality of the evaluation report as an example of "grey literature" is as good as the processes, tools and conduct applied throughout the study. It therefore makes a significant contribution towards helping readers judge the quality of what they read. Could such a system be generalised? Given the wide variety of grey

situation it has to be adapted to your actual context. This necessary adaptation is better expressed by using the term of «good practice».

9 Organsiation for economic co-operation and development OECD, http://www.oecd.org/department/0,3355,en_2649_34381_1_1_1_1_1,00.html [site visited 04.08.09]

literature it is of course not feasible and nor the aim to develop a universal system for all producers of grey literature. But a basic set of steps for guiding the production of quality output in the field of grey literature could go part way towards improving its overall quality.

References

FOPH Swiss Federal Office of Public Health. Guidelines for Health Programme & Project Evaluation Planning. 1997. Berne, FOPH. URL: http://www.health-evaluation.admin.ch (>tools >guidelines) [site visited 04.08.09]

FOPH Swiss Federal Office of Public Health. Glossary of Evaluation Terms. 2005. Berne, FOPH. URL: http://www.health-evaluation.admin.ch (>tools >glossary) [site visited 04.08.09]

Fornerod, Serge. A quoi et à qui servent les évaluations ? 2001. IDHEAP.

GLISC. Guidelines for the production of scientific and technical reports: how to write and distribute grey literature. Version 1.1. 2007. GLISC Grey Literature International Steering Committee.

Jacob, S. and Varone, F. L'institutionnalisation de l'évaluation des politiques publiques : expériences étrangères et scénarios pour la Bélgique. 2002. Louvain-la-Neuve, Université catholique de Louvain.

Läubli Loud,M. 2004, Setting Standards and Providing Guidelines. The Means Toward What End?, Evaluation 10, 237-245.

Perrin,B., 2006, How Evaluation Can Help Make Knowledge Management Real., in: R.C.Rist and N.Stame, eds., From Studies to Streams: Managing Evaluative Systems. (New Brunswick, NJ).

Widmer, Thomas. Leitfaden für Wirksamkeitsüberprüfungen beim Bund. Instrument zur Qualitätssicherung gestützt auf die Evaluationsstandards der Schweizerischen Evaluationsgesellschaft (SEVAL-Standards). Erarbeitet im Auftrag des Bundesamts für Justiz. 2005. Berne, Bundesamt für Justiz.

Widmer, Thomas, Landert, Charles, and Bachmann, Nicole. Evaluation Standards of the Swiss Evaluation Society (SEVAL Standards). 2000. Berne, SEVAL. URL: http://www.seval.ch >standards [site visited 04.08.09]

Widmer, Thomas, Rüegg, E., and Neuenschwander, Peter. Stand und Aussichten der Evaluation beim Bund. 2001. Zürich, University of Zurich.

Chapter 3

Grey Literature produced and published by Universities: A Case for ETDs

Primož Južnič, University of Ljubljana, Slovenia

3.1 Introduction

Universities and other institutions of higher education are important producers of grey literature (GL). Most of the education process at universities is based on various written essays and other assignments. This process is usually completed with some sort of written thesis or dissertation, which shows that a graduate is capable of research work and has a proper knowledge of the field.

A thesis is a written text representing the independent research and authorship of a single individual. Its purpose in higher education remains the same today as it has been for centuries, across countries and disciplines. It would be beyond the scope of this paper but still worth mentioning that this remains the principle despite various critiques of both the romantic notion of authorship and the epistemological assumptions that form traditional notions of independent scientific and scholarly research: research today involves teamwork, multi-authorship is the rule in most scientific disciplines, but the thesis remains the last bastion of single authorship.

What is surprising is the similarity of the systems that award degrees on the basis of written substantiation all over the world. Guides on how to write final student work can be applied just about everywhere. Students can follow such guidelines, even if written for American students, as an in-depth and comprehensive guide to the process of writing a thesis anywhere in the world. Even the subordinate and intermediate goals and models prepared for each stage of the process that leads to writing a thesis can be applied in the same manner. Four essentials:

1. A clear understanding of the meaning and purpose of the student research work;
2. An accurate knowledge of what constitutes an acceptable thesis;
3. A detailed plan of action;
4. A technical plan to implement research skills (Mauch and Parks 2003).

Within each essential, which can be also seen as a step toward the final goal, the students are supplied with all the tools and detailed instructions necessary for the successful completion of a thesis. These are provided by the academic mentor or

supervisor, and students are assisted by other academic staff, professors, teaching assistants, and librarians during their work. Along the way, students acquire skills and techniques that can help them cope effectively with research work and its reporting. A thesis is not a goal in itself, but rather a way to achieve certain skills and competence.

What happens afterwards? A student successfully completing a degree on the basis of a thesis receives his diploma as confirmation that he/she is ready to join the social division of work or the labour market in a certain role. The proof of this readiness, the thesis, remains at the academic institution. Traditionally, theses were regarded as library material because they were available through academic libraries. Libraries made them part of their collections, catalogued them, shelved them, and made them available to users. They were typical GL material as it was not easy to find and access it. Libraries had also an archival role since often only one copy of a thesis existed. These papers are not the only grey literature originating at universities—research contributes its share—but form by far the greatest part of it and are the most recognizable for any institutions of higher education. Their vast numbers place universities among the greatest sources of grey literature.

Some of the theses, or at least parts of them, find their way into non-grey literature, journal articles, or published congress materials and books. However, the majority and their content remain as a part of grey literature, with all the corresponding obstacles for users relative to its availability and usability. One of the defining characteristics of grey literature is that is hard to find, and theses generally fall into this category. Libraries and librarians have made some remarkable efforts to make these resources available to users. One of the best known is surely the British Library with its effort to make all British doctoral dissertations (PhDs) available from the British Library under a scheme started in 1971. But even there, the mechanisms for collecting theses have been rather chaotic and have changed over time from the ad hoc arrangement before 1970 through attempts at comprehensive collection in the 1970's and 1980's to the current situation where PhDs are obtained on demand from universities when a request is received at Boston Spa (Tillet and Newbold 2006).

Beside their formal value as a verification of acquired skills and competencies and an information source for researchers and professionals, theses have also a third purpose. They demonstrate and reflect the quality of the higher education institution where the theses were defended, an important part of any academic program evaluation. This point will be further elaborated in connection with the Electronic Thesis and Dissertation process.

3.2 Electronic Thesis and Dissertation (ETD)

The Internet has helped to solve many library and librarians' problems and relieve (academic) librarians from trivial and routine tasks. This applies to theses and

dissertations as well. The solution offered is Electronic Thesis and Dissertation (ETD). The term ETD refers to a thesis or dissertation that is archived and circulated electronically rather than archived and circulated in print. Most ETDs take the form of text uploaded in a word processing format or in Adobe's portable document format (PDF) and look very much like traditional printed theses. They reside on the Internet where they are accessible to potential users.

A major boost to ETD was the Networked Digital Library of Theses and Dissertations initiative. The Networked Digital Library of Theses and Dissertations (NDLTD) is a collaborative effort of universities around the world to promote the creation, archiving, distribution, and access of ETDs. Since its inception in 1996, over one hundred universities have joined the initiative, underscoring the importance institutions place on training their graduates in the emerging forms of digital publishing and information access (Suleman 2001). The NDLTD is an international organization dedicated to promoting the adoption, creation, use, dissemination, and preservation of electronic analogues to traditional paper-based theses and dissertations. Its website contains information about the initiative, how to set up ETD programs, how to create and locate ETDs, and current research in digital libraries related to the NDLTD and ETDs.

An overview (Edminster 2002) of these international efforts to develop a worldwide digital library of theses and dissertations focused on

(a) the need to provide developing countries with equal access to current international scholarship;
(b) the collaborative development of training materials to facilitate wider global participation in the NDLTD;
(c) the work of multi-university/library and corporate collaborations to establish centralized metadata for ETDs; and
(d) the development of multi-language search interfaces.

However, the objectives of the NDLTD were originally seen more broadly, including

- to improve graduate education by allowing students to produce electronic documents, use digital libraries, and understand issues in publishing;
- to increase the availability of student research for scholars and to preserve it electronically;
- to lower the cost of submitting and handling theses and dissertations;
- to empower students to convey a richer message through the use of multimedia and hypermedia technologies;
- to empower universities to unlock their information resources; and
- to advance digital library technology (Suleman 2001)

To gain an overview of activities relating to ETDs internationally, the web sites of every member of the NDLTD were examined. A study of approximately two hundred sites revealed that only a small percentage of the NDLTD institutions dealt with a large quantity of ETDs in 2002 (Copeland, 2003) The findings from the survey indicated that many universities could make better use of the guidance

notes relating to all aspects of ETD production, management, and use, so it should be seen as an initiative impacting on various national ETD systems.

Why national systems? Usually theses are seen as an important information resource because as a rule they are the result of research. We have mentioned already that part of their content finds its way into other publications (journal articles, congress papers, and books), but not all of it. This is an important element in national use, although we tend to forget that theses and dissertations serve to disseminate research information within local communities, especially within smaller countries and language environments. A survey by Stock (2008) of theses written in English showed important differences between European repositories. In the Scandinavian countries as well as in Belgium and The Netherlands, between 50% and 90% of (doctoral) theses are in English. In German universities the percentage of English theses has grown to reach 25%. This indicates the willingness in some countries to give the widest access possible to one's work through the choice of language and through the internet.

This is perhaps positive globalization, but it also has a negative effect. While English has become the international language of research, this does not mean that all other languages have become non-scientific. If theses and dissertations are not available in national languages, this will become an issue and a problem.

There are various national initiatives and surveys presenting the current state of theses and dissertation collections, their usage, problems with access, and the academic and research community's attitude toward ETD. But do national ETD systems work? They seem to suffer from the same problems that plague the international NDLTD system, at least judging by national reports.

In India an integrated system at the national level to locate and access theses has not been fully implemented. While just a few Indian universities have actually started ETD projects at the moment, the majority have the intention of starting such projects soon (Vijayakumar 2007). In recent years, South Korean university libraries have tried to improve user services and access to ETDs in several ways. However, authors blame the absence of an adequate policy and infrastructure to handle them at the national level for the fact that little practical progress has been made at individual academic libraries (Park 2007).

As reported for France, an integrated national ETD system still does not exist, the results of the government initiative seem disappointing, and the development and implementation of national software and services is progressing more slowly than planned. At the same time, a growing number of alternative, more or less successful local initiatives, academic networks, and open archives provide access to more than four thousand ETDs. The reasons for this paradoxical situation are various. So far, neither the government nor any other institution has had enough coercive or persuasive force to impose a unique model for ETDs. Perhaps this "unique model" is simply unrealistic and not adapted to the heterogeneous needs, behaviours, and traditions of France's scientific and academic communities (Paillasard 2004).

A study made in Slovenia revealed that only a minority of the higher education institutions have some form of their own ETD system, and not much more

intend to organize one in the near future. The great majority of these libraries allow their patrons and other users to access and use this part of their collections (theses and dissertations) only within the library premises. It can be understood from the information on their web sites that some libraries require a special author's permission before allowing access to the material (Juznic, 2009). This is changing fast as two central ETD systems emerge. At the University of Ljubljana, a new portal, the Digital Library of the University of Ljubljana (DIKUL – *Digitalna knjiznica Univerze v Ljubljani*) has been established employing the concept of local ETD systems (each faculty and department should have its own). Theses and dissertations are seen as one of the digital information resources students and teaching staff use (along with international and domestic e-journals, e-books, digital teaching materials, etc.). The same concept was used recently to establish the Digital library at the University of Maribor, Slovenia's second largest university (DKUM – *Digitalna Knjiznica Univerze v Mariboru*[1]).

ETDs can also be accessed through the National Union Catalogue COBISS where a link to the digital version can be added to an original catalogue input. Interestingly however, since it works well, the tradition of written theses and dissertations may be a real obstacle to moving to a more digital world. The National Union Catalogue COBISS has a long tradition in Slovenia and includes data about all theses defended on all levels of higher education in Slovenia.

The wide availability of the National Union Catalogue and COBISS is likely to encourage a shift to full-text databases of electronic theses and dissertations (ETDs). However, this can be also an obstacle at least for a certain period since many libraries and librarians might see it as a reason not to have their own ETD system, if "everything" would already be available. A further obstacle might be the extreme decentralization of academic libraries in older universities and the absence of any form of library services in newer universities and higher education institutions. Decentralization in its present form could mean a zealous opposition to any form of ETD system, and the absence of library services means that an ETD system cannot be built at all.

We can also see other examples why a well-established "traditional" system might be an obstacle to moving toward ETD. The UK is "currently behind many other countries in providing full text electronic access to theses produced in its higher education institutions," as reported by Tillet and Newbold (2006), even though it has one of the greatest collections of theses and dissertations and a well-developed system for their dissemination. However, once launched, ETD systems might develop faster and acquire higher usage in locations without such traditional systems.

It is clear that in recent years an increasing number of universities are building their own ETD systems or are at least considering to do so. Why are they important for every university? More and more ETD initiatives are connected with the electronic submission of theses and dissertations and other issues that help solve

1 http;//dikul.uni-lj.si

specific university problems, improve quality, and save time and money. These are usually followed by a series of promotion activities launched for teaching staff and students as well as for librarians. The electronic submission of ETDs looks to be the next step, which should be easy due to the widespread use of various e-teaching programs in which students already present their papers in electronic form for supervision and grading.

3.3 Advantages of an ETD system for a university

Generally speaking, five objectives for university or other higher education institution ETD systems can be named:

1. to make research reported in theses and dissertations more widely and easily available;
2. to initiate and encourage digital development;
3. to ease submission process;
4. to save space in libraries; and
5. to benefit the higher education process.

The first objective is very general and needs little explanation. An institutional repository includes a variety of materials produced by the university, not only theses and dissertations but also research reports, congress papers, and especially teaching materials. Some university institutional repositories are also being used as resources for electronic publications and e-journals published or originating at the university. This makes university institutional repositories different from other types of digital repositories. Adding other material, preferably licensed and Open Access (OA) electronic journals, databases, and other information sources make such portals powerful tools for students and teaching staff and support educational and research processes at the universities.

These portals have an important role in encouraging digital development overall, especially when students, as future professionals, learn to appreciate complex and interconnected portals. A series of promotional and educational activities for teaching staff and students should be launched by librarians and other information experts to raise awareness and boost their use.

Although they fail to substantiate their claims with data, many argue that electronic writing tools are transforming graduate education, enhancing mentoring and the shape of thesis content. A recent analysis of bibliographies from student research papers revealed what sources students used to support their research. While web sites were a definite fixture in student bibliographies, on average they were not the predominant source of information that one might expect given the current perception of student research. In the study, 55% of the bibliographies did not cite any web sites at all. This is an important finding to note, as it runs counter to the concerns of faculty (Carlson 2006). It might vary across the disciplines, but it is generally valid for the majority. One of the reasons for this might be that when students face submitting their work in the traditional printed format, they tend to

work or think traditionally about the information sources they use. Another reason might be the instability of Internet resources. A study of undergraduate students' citations of web sites had astonishing results: only 18% of the URLs cited in 1996 and only 55% of the URLS cited in 1999 led to the correct documents in 2000 (Davis 2001).

Since then, "URL decay" has been observed in many studies with very similar results, from infometrics showing that in each round of searching the character of the search results from the Internet was slightly different to documents appearing, disappearing, and changing (Bar-Ilan and Peritz 2009). So besides the obvious growth, they observed both decay (pages disappearing from the Internet) and modification. In biomedicine (Wren 2008), where the results are very similar, the majority of web-based resources cease to be available after a certain period of time or they are changed, something which should be more worrying and should alert us especially to the preservation of research work and scholarship in general. ETD can be a good solution for adapting to these phenomena, since it has been determined that URLs published by organizations tend to be more stable than others.

Generally speaking, the paper-based thesis submission process consists of three steps: production, submission, and preservation. Availability and use are primarily shaped by the paper version. Many universities are experimenting with electronic submission, which completely surpasses traditional paper forms. Bevan (2005) describes the issues involved in the introduction of mandatory submission of electronic theses at Cranfield University in the UK. McGill University in Montreal, Canada has undertaken a pilot project to test aspects of workflow, style sheets, metadata, and search functions (Park 2007a). In the pilot project, a new model for tracking the electronic file through the production, conversion, dissemination, and preservation processes was developed. The students first submit their theses in whichever of four authoring tools they prefer. After the completion of the examination process and thesis revision, the students submit two paper copies of the thesis to the Thesis Office and upload the electronic version. The supervisor reads and approves either the paper form or the electronically submitted final copy. The Thesis Office performs a content check on both versions, a paper copy of the thesis is sent to the library, and the library is notified that the content check has been completed.

The advantage of single-institution ETD systems is clear and obvious. A study of ETD system implementation at individual higher education institutions discovered that library administrators who implemented ETD repositories at different universities adapted their models to the needs of their institutions and their graduate students. ETD system administrators made decisions about implementation models and software and hardware infrastructure in terms of human and technical resource allocation (Yioris 2007). These decisions are difficult to achieve at the international or even the national level, and this gives the advantage to local systems.

The next step is seen as the electronic submission of ETDs automatically building the repositories. The permanent and secure preservation of documents is

often an issue; the tension between libraries' two-fold responsibility of preserving and providing access to information takes on particular significance with ETDs. As the examples have shown, many universities balk at the idea of allowing students to submit work exclusively in electronic form, and they continue to require what is perceived to be a more "permanent" print copy for archival purposes. As complementary to print, some universities will accept an archival version on CD-ROM, but there are concerns as to the long-term durability of this technology (Edminster 2002). This form of storage will ease the pressure on library space, where the great number of theses tends to occupy space that is often needed for other activities either of academic libraries themselves or of other academic departments.

3.4 Plagiarism

The preservation and availability of ETDs at all levels is not the only concern universities and other higher institutions have regarding them. There is also a concern regarding plagiarism and other forms of cheating. Plagiarism is the nightmare of higher education, often a theme not to be discussed in public. It is even hard to uncover the extent of it. Over a three-year period, McCabe (2006) surveyed more than 80,000 students and 12,000 faculty in the United States and Canada and confirmed that plagiarism is a significant issue. For example, if the four behaviours in which students engage least frequently - turning in work copied from another, copying large sections of text from written sources, turning in work done by another, and downloading or otherwise obtaining a paper from a term paper mill or website - are combined, it is clear that 16% of all undergraduate respondents and 8% of responding graduate students reported one or more of these behaviours in the past year. In contrast, a surprisingly large number of faculty (79%) report they have observed one or more instances of these behaviours in the last three years, driven in part by a perception that a large number of students (59%) have copied material almost word for word from a written source without citation. Due to their "grey literature" nature, ETDs are often seen as the main source of students' "cut and paste" work.

While ETDs not only improve access to grey literature, they also serve other two purposes. They have the potential to change, modernize, and improve the way students acquire their future skills and to improve the quality of higher education to which various sorts of plagiarism pose a constant threat. At first glance, this statement appears to contradict the fact that when something is in digital form and freely available it is easier to become a source of copying or plagiarism in general. The traditional paradigm was to make this material available through academic libraries. The Internet has helped to simplify this process and relieve academic librarians from trivial and routine tasks. It has also made it easier for all potential users, often students themselves, to access these materials, adding to other materials they can use that are part of GL materials. This sounds like a great leap for-

ward if current research did not indicate that academic plagiarism is now a very serious problem worldwide.

So it is easier for students to plagiarize from ETDs because of the increased access to electronic documents and simple copy and paste functions. The features of search functions, however, make detecting plagiarism easier as well. Every university has policies in place regarding plagiarism, and these must be enforced along with the proper application of fair-use guidelines (Yioris 2007). Why the second? There are many good technical methods of detecting plagiarism, but students can not be left alone in the fight to prevent it. Librarians can help considerably by educating students on how their work will be assessed and the potential traps of possible plagiarism. The difference between copyright violation and the threat of plagiarism is often confused in discussions about intellectual property. Plagiarism occurs when someone poses as the author of a work; copyright infringement occurs when someone uses another's work without proper authorization or citation. Students rarely understand the difference, and librarians have the expertise and authority to help them make the distinction.

Librarians need to get more involved in helping students write theses and dissertations and create their electronic counterparts. Active participation in the creation of theses and dissertations, the ultimate demonstration of higher education, could certainly have positive status repercussions.

In theory, librarians are seen as experts who understand user needs and perceptions. They know what works and what does not. They know how to help, inform, persuade, and teach users (Bailey 2005). They could serve as more than just "plagiarist busters," but this does require that librarians improve their own knowledge of the issues regarding academic integrity. They should be able to promote a more complex understanding of the Internet and a critical approach to research and writing. The problem is not that students today are more dishonest but that their experience—particularly with the Internet-based transfer of information—has led them to form different attitudes toward information, authorship, and plagiarism (Wood 2004). Student perceptions of what constitutes dishonesty, what cheating means, and what plagiarism is differ from those of academic staff. It would not be fair to say that this is the result of a decay in moral values, but rather often of the different experience new generations have. Generally, these generations are characterized by an increased use of and familiarity with information technologies and digital formats, which is accompanied by a different attitude toward propriety rights and copyrights. However, plagiarism in student theses still constitutes one of the most visible and also most dangerous problems.

Maybe the best example is citations. The average student regards citations as annoying details with little relevance to the work. On the other hand, academic staff understand that established citation conventions are the basis of research and scholarship and prove the validity of one's research work. Through citation, researchers acknowledge their debt to their predecessors, and they often constitute the difference between plagiarism and one's own work, showing what new has been added. They also show the students' understanding of the research process and their skills. Librarians can serve as mediators between academic staff and

students, teaching or advising students regarding their written work and reading through it for details that are connected to their work and services, for example, the proper use of various information sources. It is often stated that librarians need to shed their preconceptions about how academic staff and librarians should collaborate and accept shared responsibility for student learning (Doskatch 2003). They should be more involved in students' work and not just "behind the counter" in their libraries. Librarians should get more involved both in making theses available and at the same time in fighting plagiarism and how their expertise in dealing with different information sources, including those called "grey literature," can be used to help teaching staff in their struggle to maintain the quality of academic education. This is also one of the factors turning traditional library tasks and services toward the more professional expertise expected by information technology experts. The survey of academic libraries in Slovenia have shown that they see plagiarism as an issue and a problem but generally speaking, the librarians thought that plagiarism is the primary concern of mentors and teaching staff and not theirs (Juznic, 2009).

Technically, it can be assured that ETDs are checked for the most blatant cases of plagiarism using applicable technical methods. Many "check for plagiarism" commercial and in-house/open source programs are available and ready to use. The wider use of ETDs would make these programs more accurate since they can not check plagiarism from written theses and dissertations that are not available in electronic form or on the Internet. Of course, English is generally the language of the materials these programs check, although not exclusively, and fighting plagiarism in other languages is also making progress. The National Registry of Theses and the Plagiarism-Tracing System, a project involving twenty Czech and Slovak universities, is an interesting example. The project has two main parts: the first part gathers metadata on theses and the second, the Plagiarism-Tracing System, serves for detecting plagiarism. The system will facilitate academic staff in discovering possible cases of plagiarism.(Pejsová and Pfeiferová 2008)

The electronic submission of ETDs will be next step, which should be easy due to the widespread use of various e-teaching programs in which students already present their papers in electronic form for supervision and grading. There are other problems in addition to the technical ones. The latest survey in Slovenia asked students for their views on ETDs (Zeleznik and Juznic, in press) to determine if they are prepared for this step. The hypothesis was that students would have no objections since their use of the Internet is a well known fact. The survey of students at all four public universities in Slovenia showed that they want wider access to theses and dissertations and that they want universities to provide them in ETD form. A large majority agreed that the written work of their predecessors is an important information source and that there should be a possibility of making it more available. Students are mainly interested in certain parts (chapters) and, interestingly enough, in the bibliography or reference section in order to inform themselves about relevant related literature. One group of students uses theses like any other research material, for the sake of content, while other group

sees theses as a good substitute for searching through various bibliographic databases. Some 90% of all the surveyed students would use ETDs more if they were available in electronic form on the Internet.

On the other hand, half of the students do not think having their theses only in electronic form without a printed counterpart is a good idea. Furthermore, only half of the surveyed students think that having all theses (including their own) freely available to everyone is a good idea, and the same number think that this would allow someone to copy their work and claim it as their own. A substantial percentage of the surveyed students have very serious doubts about ETDs and the exclusively electronic submission of their theses.

Perhaps an update of this survey in one or two years could provide more evidence on the relationship between the availability of theses and the students' positive attitude toward ETDs. It would also be interesting to compare our data with that from other universities or other institutions of higher education.

3.5 Continuing challenges and future developments

The electronic submission of ETDs must be the next step. Many academic libraries might think various issues are an obstacle to creating ETD systems, including the risk of plagiarism and the lack of funding, administrative support, and regulation. However, those that have already started creating their own ETD systems should prove them wrong and demonstrate the possibility that the infrastructure, technical expertise, and financial support to create ETD systems already exist in their own institutions. Effective awareness programs are required to increase their visibility and emphasize their usefulness. The complete electronic submission of theses and dissertations can be the decisive point toward implementing ETD systems and is therefore worth special effort and investment.

We also need other activities to promote the concept of ETD systems. According to current data, workshops and web documents are most often used to educate students about ETDs, although faculty and administrators learn about them mainly through presentations, lectures, and seminars. The methods might be different in different environments, but the fact is that approaches must be different for different users. Even if ETD systems benefit students, professors, and the public alike by enhancing graduate education, expanding graduate research, and increasing a university's output quality, the activities must be tailored for the different audiences. Universities need to recognize the potential value of accessible ETDs since theses and dissertations reflect an institution's ability to lead students and support original work. An interesting observation is that when ETDs are in an accessible place, students and teaching staff will make judgments regarding the quality of a university by reviewing its digital library. Universities must respond accordingly, ensuring they provide the resources and training students need to incorporate new literacy tools such as animation, graphics, sound, and streaming multimedia (Edminster 2002).

This may be seen today as a distant future, The uncertainty created by the relatively recent introduction of ETD systems and the absence of national policies and frameworks in this area hinder their rapid adoption. What we might need is an ETD submission protocol, implemented and tested for different institutions. As a result of the different ETD projects, recommendations can be made and different approaches can be decided on. It will be exciting to see something regarded as a grey literature in the past and treated accordingly, become the core of higher education activities and a centerpiece of a university's reputation.

Librarians are getting involved both in making materials available and at the same time in fighting plagiarism and how their expertise in dealing with different information sources, including those called "grey literature," can be used to help teaching staff in their struggle to maintain and improve the quality of academic education. This is also one of the factors turning traditional library tasks and services toward the more professional expertise expected by information technology experts.

3.6 Literature

Bailey CW. (2005), The role of reference librarians in institutional repositories. Reference Services Review. 33 (3), pp. 259-267

Bar-Ilan, J., Peritz. B. (2009) The lifespan of "informetrics" on the Web: An eight year study (1998–2006). Scientometrics, 79 .(1), pp. 7–25

Bevan, S. (2005), "Electronic thesis development at Cranfield University", Program: Electronic library and information systems, 39 (2), pp. 100-11.

Carlson, J. (2006), An Examination of Undergraduate Student Citation Behavior The Journal of Academic Librarianship, 32 (1), pp. 14–22

Copeland, S., Penman A., Milne R. (2005), Electronic theses: the turning point. Program: electronic library and information systems. 39 (3), pp. 185-197

Davis PM., Cohen SA., (2001), The Effect of the Web on Undergraduate Citation Behavior 1006-1999, JASIS 52 (4), pp. 309-14

Doskatch, I. (2003), Perception and perplexities of the faculty-librarian partnership: An Australian perspective. *Reference Services Review,* 31 (2), pp. 111-121

Edminster, J., Moxley, J., (2002), Graduate Education and the Evolving genre of Electronic Theses and Dissertations. Computers and Composition 19, pp. 89–104

Juznic, P. (2009)., Grey Literature Produced and Made Available by Universities – Helping Future Scholars or Plagiarists? The Grey Journal, 5 (1), pp. 23-30

Mauch, JE., Park, N. (2003), Guide to the successful thesis and dissertation: A handbook for students and faculty. New York : M. Dekker, Boulder

McCabe DL. (2006), Cheating among college and university students: A North American perspective. international journal for educational integrity 1 (1), pp 1-11

Paillassard, P., Schöpfel, J. & Stock, C.(2004), How to get a French doctoral thesis, especially when you aren't French. GL6 conference. Available http://www.greynet.org/images/GL6,_Page_143.pdf

Park EG, Nam Y, Singe OS. (2007), Integrated Framework for Electronic Theses and Dissertations in Korean Contexts. The Journal of Academic Librarianship, 33, (3), pp. 338–346

Park, EG., Zoo, G., McKnight, D., (2007), Electronic thesis initiative: pilot project of McGill University, Montreal. Program: electronic library and information systems. 41 (1).

Pejšová, P., Pfeiferová. M. (2008) Grey Literature in the Czech Republic. GL10 conference, Available: http://www.greynet.org/images/GL10,_page_21.pdf

Stock C. (2008), Open access to full text and ETDs in Europe: improving accessibility through the choice of language? The Grey Journal, 4 (2), pp. 73-82

Suleman H et al, (2001) 'Networked Digital Library of Theses and Dissertations: Bridging the Gaps for Global Access—Part1: Mission and Progress,' D-Lib Magazine, 7 Online. Available: http://www.dlib.org/dlib/september01/suleman/09suleman-pt1.html

Tillett, S., Newbold, E. (2006), Grey literature at The British Library: revealing a hidden resource. Interlending & Document Supply 34 (2), pp.70–73

Vijayakumar, J.K., Murthy, T.A.V., Khan M.T.M. (2007), Electronic Theses and Dissertations and Academia: A Preliminary Study From India. The Journal of Academic Librarianship, 33 (3), pp. 17–421

Wood, G., (2004) Academic Original Sin: Plagiarism, the Internet, and Librarians. The Journal of Academic Librarianship, 30 (3), pp. 237–242

Wren JD (2008), URL decay in MEDLINE—a 4-year follow-up study Bioinformatics, 24 (11), pp.1381-1385

Yiotis, K., (2008), Electronic theses and dissertation (ETD) repositories What are they? Where do they come from? How do they work? OCLC Systems & Services: International digital library Perspectives, 24 (2), pp. 101-115

Part I, Section Two

Collecting and Processing Grey Literature

Owing to its ephemeral or fugitive like reputation, collecting grey literature remains a challenge to library and information science professionals. Grey items such as reports, proceedings, or working papers cannot be purchased or bought like journals and books. There is no special agency or supplier for grey materials.

Buying information nevertheless is part of the traditional library role, together with gateway and archive functions[1]. In line with the economic definition of grey literature, "material that usually is available through specialized channels and may not enter normal channels (…) of (…) distribution", one comes to understand that a grey collection calls on specific attention, competency, and procedures.

The British Library has an outstanding experience with collection and processing grey literature, especially regarding conference proceedings and Ph.D. theses as well as technical and scientific reports. The first chapter offers firsthand knowledge of academic holdings and the specific place grey items occupy. Newbold and Grimshaw's first observation is that "the representation of grey literature in library collections varies considerably. In some specialist and technical libraries the majority of the collection may consist of grey material, while in other institutions it may be [only] a small percentage of the total holdings." The authors add that "librarians have traditionally been wary of grey literature, due to the difficulties involved in identifying, acquiring, cataloguing and, shelving it." They are keen to note that one of the most common words that comes up in conversation with librarians about grey literature is 'difficult'. The chapter proceeds as a kind of compact manual for librarians in charge of grey collections, especially in digital format.

The second chapter in this section helps in understanding the relation between the production and collection of grey literature on an academic campus. Siegel starts with empirical evidence on her own Portland campus. She asserts "institutional grey literature was being and had been produced on campus for quite a long time. The library holdings included an assortment of these reports (…). There was no coordinated effort for the collection of these reports." She then reviews other American and European initiatives on digital grey materials in the emerging infrastructure of institutional repositories - underscoring the role academic libraries (must) play in order to improve bibliographic control of this very specific stuff.

---

1 B. Heterick & R. C. Schonfeld (2004). `The future ain't what it used to be'. *Serials: The Journal for the Serials Community* **17**(3):225-230.

She then challenges “to optimize discovery - interoperability should be a key factor in determining whether to ‘locate’ grey literature in the library catalog, an institutional repository, or both”. In a way, Siegel’s analysis is a reply to Roosendaal’s conceptual study on value chain and business models applied to the campus.

This section’s final chapter “reviews recent legislative and case developments in the area of copyright law affecting the collection, preservation, including digitization and dissemination of grey literature.” Lipinski examines a number of frameworks among which include: information policy related to copyright in a “grey” context; Section 108 (fair use) in library and archive reproduction and distribution; the orphan works; and threats to the public domain having repercussions for grey literature. He finishes the chapter with a rich body of notes, comments, and references. In following this line of discourse, Lipinski insinuates that even if the US environment is different from that of the EU - national and international legal frameworks on intellectual property are converging.

This section no doubt conveys a more “traditional” library oriented perspective to the monograph, which brings us to the guiding questions posed to the reader:

How can one best define and organize specific grey collections? Would it be through document categories, disciplines, distribution channels, producing bodies, or a mix? What is the specific impact of digital resources on the acquisition policy of academic libraries for grey literature? And, from a legal perspective, in what way does the acquisition and processing of grey material differ from that of journal and book collections?

Chapter 4

Collection building with special Regards to Report Literature

Elizabeth Newbold and Jennie Grimshaw
The British Library, UK

4.1 Collection Building

In this chapter we will look at the meaning of collection building in the digital age, concentrating on some of the drivers and issues that can be applied to any type of library, but with an emphasis on our experiences in the United Kingdom. Specific to this chapter, we will look at report literature in the sciences and social sciences. This can include, but is not limited to: research, practice, evaluation and development reports distributed by government, international or intergovernmental organisations; policy, regulatory and guidance materials from central or local government and agencies; reports or technical papers produced by research institutes, think tanks or consultants; and material from voluntary and community sector organisations and charities. Firstly, let us consider what we mean by collection building and collection development?

Collection building is a process usually conducted over a period of time that shapes a collection of resources into a cohesive, balanced and useful set of material for a given user community. Collection building encompasses not only material owned by a library in both physical and electronic formats but the also fee-based electronic resources to which it subscribes and free Internet sites to which it links. Collection building takes place not only in all academic, research and special libraries but also among providers of abstracting and indexing databases and information gateways. It includes the organisation and presentation as well as the acquisition of material.

4.1.1 Why build collections?

Mostly librarians have in the past built collections to have materials available for patrons 'just in case' they are needed. However, is this still a relevant and useful activity in an increasingly digital and connected environment where information is but a click away on any networked computer? Should librarians still be building

collections and what is the purpose of this activity? A collection should be a useful aggregation of resources based on user needs and demands. Easier access to material online has placed users at the centre of collection building activities and libraries are adopting more 'just in time' approaches to collection development. Collection building is not just about what a library or information unit holds in either physical or digital locations but also about how it enables resource discovery and facilitates access to material. The role of collection development and collection building is to organise and index the wealth of information available in print and electronic formats so that users can home in on what they need quickly and accurately and to ensure that access is maintained over time. So, although collections are now hybrids including both print and electronic material, building them requires traditional activities and skills:

Selection

Selection has always been fundamental to collection building. Selection brings together material from different sources and locations and in a range of formats to provide a coherent resource for users. While researchers have domain expertise in their subjects, the role of the librarian is to have domain expertise in the information related to those subjects, where to find it, what quality criteria to apply and how to source it. The role of selection goes beyond that which is found by a simple internet search; it identifies a range of sources of relevant material in print and electronic formats, monitors their ongoing availability and scans the horizon for new sources.

Organisation

One of the most important aspects of collection development is the organisation of information so that it can be quickly, comprehensively and accurately retrieved by users. Report literature is found in multiple formats and is produced by a wide range of organisations. It may be owned by the library or information unit, have been deposited in an institutional or subject repository, be included in a fee-based database, or be published free on the Internet. Traditional library skills of resource description and indexing are needed to facilitate its efficient discovery by users.

Long term access

One of the reasons for building collections is to ensure long term access to material. Grey literature in hard copy has always been difficult to find and acquire because it is not covered by mainstream abstracting and indexing services, is produced in limited print runs, and is not available through the book trade. With the advent of web publishing grey literature is in some ways easier to discover through simple internet searches, but long term access to the material is much less

certain. Much of the material is now only available as a web page or as a document attached to a web page and has no print equivalent. Web pages are transient in nature; some estimates put the life span of a web page at just 75 days (Lawrence, 2001). Links to material on web sites are often broken when these are redesigned and uniform resource locators (URLs) are changed. Whether the content has disappeared or only moved is to some extent immaterial; librarians building collections will need to solve the problem of ensuring long-term access to it.

Continuity of access - switch from print to online only publication

Many reports and especially official publications texts are published in series. In some instances, especially in the case of government publications, the series may have been in existence for decades. Libraries hold long runs of the print publications, which can suddenly cease when the documents migrate to the web and the hard copy version is no longer produced[1]. For the convenience of researchers seeking information over time, the successor electronic versions need to be collected and made accessible by the same institutions. Ideally the whole time series should be made available at one location, and the print and electronic versions linked in library catalogues.

4.1.2 What is the role of grey literature in a collection?

As we know there is considerable debate over the definition of grey literature. The representation of grey literature in library collections varies considerably: in some specialist and technical libraries the majority of the collection may consist of grey material, while in other institutions it may be a small percentage of the total holdings. Librarians have traditionally been wary of grey literature, due to the difficulties involved in identifying, acquiring, cataloguing and shelving it. One of the most common words that comes up in conversation with librarians about grey literature is “difficult”. Grey literature is an often-overlooked resource and does not always figure in the collection development policies or selection guidelines of libraries (Lehman and Webster, 2005).

The fact that grey literature has often been overlooked and is difficult to deal with does not diminish its worth as it offers many benefits to users. Grey literature, especially report literature, provides access to high caliber research often not published elsewhere. Documents produced and prepared for one organisation or purpose may have relevance or resonance for other audiences. It is

1 For example UK Civil Service Statistics have been produced since 1950 and are available online at http://www.civilservice.gov.uk from 1970. They are no longer produced in print and the data is now produced and provided by the Office for National Statistics at https://www.statistics.gov.uk The run of the series is now split between print only editions, dual print and electronic versions and at least two different web locations

often free or low cost and can provide a faster route to publication than that offered by academic monographs or peer-reviewed scholarly journals. Information in research reports is often more detailed than in journal articles, where page limits may have been imposed. Moreover they cover failures as well as successes.

There is a gradual movement among researchers to recognize the value of grey literature and it is seemingly becoming an increasingly important and useful resource, particularly for multi-disciplinary research and in areas such as systematic reviews in evidence based health and social policy, in the dissemination of best practice in social care, and in the relatively new field of systematic reviews in environmental sciences[2]. Systematic reviewers in all fields not only need access to research results published in report literature, but also themselves publish their work as reports. They are thus both users and producers of grey literature. The growing importance of evidence based practice in applied sciences such as healthcare and in social policy is creating greater awareness among researchers of the relevance and usefulness of grey material. This is in turn generating a demand for its inclusion in library collections.

4.2 Approaches to collection building

As grey literature is in essence not commercially published it has to be obtained in most cases directly from the producers. In the 'traditional' print environment the two main mechanisms for acquiring material were either purchase on subscription or as individual items, or by donation. Both of these approaches are still relevant today but given the migration of grey material to web-only publication libraries are faced with a slightly more complicated range of options for collecting it:

- reliance on internet access – instead of libraries spending time creating, organising and providing access to a physical collection, the role of the information professional is to assist researchers in locating material on the web.
- downloading and archiving – including the harvesting of metadata, downloading of individual documents from the internet, and the archiving of whole web sites
- use of institutional and discipline repositories
- use of commercial and non-commercial services

It is highly likely that most organisation and information professionals adding grey literature to their collections will need to consider all approaches as at the time of writing, material is available in many formats and has not migrated fully to the digital world.

2 Further information on systematic review in environmental sciences is available from the Collaboration for Environmental Evidence http://www.environmentalevidence.org/

4.2.1 Reliance on the internet

Many researchers argue that all the materials that they need are available free on the Internet and accessible via search engines (Snyder, 2008), and there is a growing body of research on user behaviour that illustrates the use of Google and other search engines as the premier tools for information retrieval. Libraries need to make a decision on whether to continue to seek to own material or to take a linking approach to building a collection of electronic materials. If a linking approach is chosen, the librarian or information professional needs to consider the stability and longevity of the source. They then need to consider how to provide the link to the resource, for example via a portal, through home-produced subject guides or by links from records in a traditional library catalogue. In this scenario the role of the library has completely changed. Instead of creating, organizing and providing access to a physical collection of locally held material, the information professional's role is to assist researchers in locating the materials they require via a range of free and priced electronic resources and to provide links to material held elsewhere.

This route has been followed by a number of libraries to a greater or lesser extent, with some information professionals providing sophisticated linkages to content whilst others are assuming more of an enabling role, training end users in electronic information retrieval techniques rather than building collections of material. In principle, the availability of such a wealth of information online should make research much quicker and easier, and negate the need to own material. However this approach may be short sighted for a number of reasons. There is ample evidence that the presence of grey literature on the internet is ephemeral due to organisational change, web site redesign and the removal of material considered to be out of date, leading to broken links, loss of access and very frustrated users.

Problems of loss of access

When organisations change name, merge or are abolished material on their websites is often not migrated to the site of the new body. The old site may be maintained for a limited period, but not indefinitely. The skill for the information professional is to anticipate when this may happen and react accordingly. Whilst this may sound like crystal ball gazing, such proactive monitoring and awareness of the changing information landscape is becoming an essential skill for collection building and ensuring continuing access to material. Whilst some changes to web sites and loss of material may happen without any prior warning, others are predictable, for example the general overhaul of public communication which follows a general election when a different political party comes into power.

Other examples of organisational reforms leading to radical web site redesign include changes in the machinery of government, when departments split and merge. Departmental histories in the UK are extremely convoluted and it is not

unusual for departments to split, merge, change name or disappear within the lifetime of a government and not just due to changes following a general election. With each change the web site is redesigned and it is not unusual to find that some information has been deleted during the transition. Other information will have moved to a new web address. In these situations there is usually some warning but it is essential for the librarian to react swiftly and it requires effort on the part of the library to ensure that key documents are tracked down and links reestablished. National web site archiving programmes may help. In the UK the UK Government Web Archive and Web Continuity Project developed by The National Archives seeks to ensure that links to government websites continue to work and that pieces of online information cited remains accessible in perpetuity[3]. Initiatives such as this ensure that material remains available and enables librarians to adopt a linking approach to collection development with more confidence.

Information may also be deleted simply because it is out of date or no longer reflects current policy. It would appear to an external audience that some organisations have prioritized website housekeeping without regard for the value of the information on the site and take the approach of simply deleting older information. The length of time a webpage or website has existed is not an adequate indicator of value and this approach causes immense difficulties to librarians as they attempt to maintain links to older information. Automated tools that check website links are a useful aid, but librarians need to develop in depth intelligence on the practices of key organisations in order to anticipate changes and take appropriate action. In other cases material remains online but is moved and the URL is changed, so that links to pages and documents bookmarked by users, cited in academic research, or retrieved by search engines are broken, again requiring the expenditure of scarce resources to track down the new location.

4.2.2 Downloading and Archiving

Having considered reliance on the internet, what are the options if the host site is not considered stable enough for long-term maintenance of access? We shall now consider the feasibility of downloading or harvesting individual documents or archiving whole websites. An attractive option is the harvesting or manual downloading of individual documents freely available on the internet.

Individual documents

Libraries wishing to download and keep individual documents in their subject area published free on the Internet have two options. They may choose to seek formal, written permission to download and archive from the web site publisher and/or the

3 Further information on the UK Government Web Archive and Web Continuity project is available at http://nationalarchives.gov.uk/webarchive/

rights holder, or they may opt for an informal verbal assurance that there is no objection. Experience at the British Library (BL) suggests that the latter approach would be more fruitful than the former. An action research experiment took place within Social Sciences Collections and Research at the British Library in April-July 2008. Staff contacted a sample of UK organisations in the field by telephone or email and initiated a dialogue aimed at persuading the body to voluntarily deposit their electronic publications for long term preservation and use within the reading rooms. This piece of action research identified a number of barriers:

- Finding an appropriate contact, especially one that had the authority to agree to electronic deposit, proved challenging. Organisational structures and staff changes added to problems here, with the final agreement often needed from senior management representatives.
- Organisations were very wary of e-deposit and were unclear about who had power of decision to authorise it. They were unclear about what they were "letting themselves in for" and were fearful of unforeseen consequences.
- Legal barriers proved a deterrent, especially the burden placed on publishers by the Library's requirement that they get permission to deposit from third party rights holders, i.e. authors of individual reports. This involved them in a great deal of work. On the other hand, many reacted with puzzlement and asked why the BL did not simply download a copy of whatever free web documents it required, referring to the example of the activities of the Internet Archive as a precedent. In other words, they were quite happy for us to gather the material, provided that they did not have to take the risk of signing a formal licence or use scarce resources contacting authors of individual reports to get their permission to deposit.
- Lack of interest within the organization was a significant problem, when staff was busy and the question of deposit of electronic publications did not figure on their list of important things to do. In the case of small charities, staff, often volunteers, are focused on serving their users and do not have spare time or energy to engage with electronic deposit issues.
- Technical barriers also manifested themselves, as not all organisations had staff with sufficient knowledge or competence to manage the transfer if a "push" rather than "pull" approach was used.

The downloading of individual documents for retention in library collections is something of a legal minefield even though many organisations offer the option of downloading documents from their websites free of charge at least for private research and study. Information professionals need to be aware of the intellectual property rights associated with that content, which are often complicated and time consuming to unravel. In some instances it is perhaps easier to fall back on to the familiarity of print copies of documents, where they are available, to avoid the added challenges associated with grey literature living wild on the web.

Web site archiving and harvesting

Another approach to the securing of long term access to grey literature published on the web is through web site archiving programmes. In the UK the British Library has been selectively archiving sites since 2005 initially as part of the UK Web Archiving Consortium (UKWAC)[4].and since 2008 as an individual contributor to the UK Web Archive. Tight legal constraints prevent the archiving of web sites without permission from the rights holders. In order to comply with the letter of UK copyright law, written permission to archive needs to be obtained from the web site publishers and other rights holders, including authors of individual documents and even contributors to discussion forums. Sites which the BL and its partners have gathered are made publicly accessible via the UK Web Archive web site[5], permission therefore has to be sought to both gather the material and make it publicly available. Seeking these permissions has proved time consuming and problematic and the success rate is currently very low at about 25% of requests made being granted. Common concerns raised by organisations from which permission to archive is sought are similar to those indentified in the experimental programme for requesting authorisation to download individual documents and include:

- Uncertainty about who in the organisation has authority to give permission, and lack of time and motivation to address the issue.
- The burden laid on the site publisher of gaining permission for the BL to archive material on their web site where the copyright is held by a third party. Such material includes documents like research reports where the copyright is retained by the author, contributions to discussion forums and blogs, and images licensed from picture agencies. It is often impractical for website managers to look through large sites for third party copyright material and then to negotiate with rights holders.
- Concerns about privacy issues, when personal information about individuals is included on web sites

Thus the barriers to collection building in the digital age using materials available free on the internet are perhaps as much political or legal as technical. We have used examples from our experiences in the UK, but programmes in other countries

4 The consortium was formed in 2004 by six institutions working in partnership: the British Library, the National Libraries of Scotland and Wales, the Wellcome Trust, the National Archives and the Joint Information Services Committee (JISC) to select and archive web sites. It is now evolving into a forum for policy development and sharing of technical expertise and best practice, with an additional advisory function in support of any institutions interested in developing a web archiving programme. In early 2008 the BL launched its new Web Curator Tool, (WCT) software developed in-house for archiving web sites. Access to the WCT, and to a hosting service for archived web sites was made available on a subscription basis, initially to the original UKWAC partner.

5 UK Web Archive is available at http://www.webarchive.org.uk

face similar if slightly different problems[6]. Even so, web site archiving provides a useful tool and option for collection building. There is a growing use of web archiving tools to build collections of web resources for different user communities. It is hoped that, as web archiving activities continue to grow, the problem of the requirement to get formal permission to archive from rights holders will be resolved through new legislation, at least as far as the British Library is concerned.

4.2.3 Use of Institutional Repositories and Discipline Repositories

Collection building may also be undertaken internally as a mechanism for storing and preserving the research outputs of organisations. Many research organisations are setting up institutional repositories and an increasing, amount of data is being deposited in institutional and discipline focused repositories.

Much of the activity has been in higher education institutions or research intensive organisations and most of the data collected is in the form of academic journal articles or theses, but it can include reports, conference presentations, working papers, statistical datasets, visual media, and sound recordings. Institutional repositories offer a hub for collecting and preserving the intellectual output of an institution in digital form. They are usually, but not always, open access, and their general aim is to make publicly funded research available, to showcase the research outputs of university departments, and to preserve them for the lifetime of the repository. Content can be retrieved via a number of routes including links from the Library catalogue, search engines such as Google, the repository's own search interface and specialist services such as Intute Repository Search (IRS)[7]. It should be noted however, that the quality of the search results is dependent on the quality of the metadata in the original repositories (Kerr, 2008).

There is evidence that digital repositories with high quality structured metadata which is accessible to search engines are successful in making content accessible to users outside of the host institution. Research at Oregon State University showed that only 25% of users of the digital repository were local. The remainder, were users external to the university from other parts of the USA, Canada, India, Europe and Asia including the Middle East. Twice as many users from Europe and Asia access Oregon University's digital library as from the USA. These external users access the digital library via search engines or the Integrated Library System (Reese, 2008).

---

6 Jacobsen, G. Web Archiving: Issues and Problems in Collection Building and Access. Liber Quaterly 18(3/4), December 2008 pp 336 – 376

7 The IRS project is a continuation and enhancement of work carried out under the ePrints UK Project, which aimed to harvest metadata from institutional and subject based eprints archives using the Open Archive Initiative Protocol for Metadata Harvesting. The IRS project builds on this initial work and in the Summer of 2008 was searching across 321,038 working papers, journal articles, reports, conference papers and other scholarly digital objects deposited into 89 UK eprints repositories.

When populating an institutional repository, managers and authors need to comply with intellectual property rights. This will involve identifying the rights holder, and if this is an entity other than the author, e.g. a commercial publisher, gaining their permission to place a copy of the work in the repository. Unfortunately this situation is complicated because authors either do not know who owns the copyright in their published works or are unaware of the limitations that agreements they have signed with commercial publishers put on their re-use. For example, Nieminien (2008) reports that authors at Bradford University were surprised to learn that, while publishers may be willing to permit self-archiving in open access repositories, they are not keen to allow the final published PDF versions of works to be placed in institutional repositories. There is a clear need for further training and awareness building in this area, for authors to be encouraged to retain copyright in their scholarly output (Oppenheim, 2008) and for librarians and information professionals to understand how this material can be re-used. As it is, accessibility outside of the host institution makes institutional repositories an increasingly useful tool for improving and widening access to grey literature.

4.2.4 Commercial and Non-Commercial Services

Commercial services

So far we have looked at options for acquiring or accessing grey literature which is available free of charge, although there are always costs associated with the selection and maintenance of access to the material. We shall now briefly consider some of the commercial services available.

One of the most common forms of dissemination of report literature used to be microform. Although this is still a distribution channel for some producers, it is becoming less commonly available and the most prevalent form of dissemination, and preferred choice for access, is now online internet based services.

Some commercial services are emerging which aggregate and provide fee-based access to full text grey literature. For example, PsycEXTRA[8] produced by the American Psychological Association, is a companion to the scholarly PsycINFO[9]. Most of the content was written for professionals and distributed outside of peer-reviewed journals. It includes full text of research reports, newsletters, policy statements, factsheets, annual reports and consumer brochures. In 2008 CABI made the decision to add full text to the standard subscription to the CAB Abstracts[10] database, which includes conference proceedings and reports from government and international organisations. It aims to be a permanent sustainable repository and grow by 10,000 documents per year (CABI, 2008). In

8 PsycEXTRA information available at http://www.apa.org/psycextra

9 PscyINFO information available at http://www.apa.org/psycinfo

10 Information on CAB Abstracts available at http://www.cabi.org

this instance grey literature is discoverable alongside mainstream journal literature to satisfy growing user expectations of immediate access full text documents. This approach is similar to PsycEXTRA in that they both aim to be a repository for full text materials, but they differ in that one is concerned solely with grey documents and the other provides grey material alongside conventional literature. These two similar but different approaches perhaps reflect differences in how researchers in various disciplines approach information seeking and use of grey literature. However it has to be said that this is still relatively unusual and most indexing services simply link to full text documents on the web, sharing the ever present problems of links which break and content which has been removed. Increasingly we are seeing abstracting and indexing services providing access to full text documents but they still predominantly focus on peer-reviewed journal articles and not grey literature.

More common are bibliographic databases that index report literature. Two long running and well known such services are the National Technical Information Services (NTIS)[11] database which indexes reports on research sponsored by the United States and selected foreign governments and the Education Resources Information Center (ERIC)[12] database which covers both journal and report literature in the field. Both these services provide bibliographic records and are backed up by an ordering mechanism for documents. Services such as these are constantly evolving and they too are starting to offer a growing collection of full text content.

The obvious constraint on use of commercial services for building collections will be the size of acquisitions budgets but they can be a cost effective and reliable option. It is perhaps more common to assess them using criteria laid down by an electronic resources collection strategy, but for databases containing full text it is important to also consider them as part of the overall collection of primary research materials. A database such as PsycEXTRA provides a valuable resource in psychology as it offers discovery and delivery of full text through a single stable platform. One of the main barriers to wider use of grey literature is the difficulties researchers can face in getting hold of the document once a reference has been found. Commercial services which aggregate hard-to-find materials in full text have a valuable role to play in enabling swift access. While there is cost associated with subscription to commercial resources, when this is compared to the life cycle costs associated with selecting, acquiring, processing and storing individual documents and the time consuming and resource intensive activities associated with maintaining links to free internet versions, they start to become very attractive options.

11 NTIS National Technical Information Service http://www.ntis.gov

12 ERIC Education Resources Information Services http://www.eric.ed.gov/

Non commercial services

In addition to subscription based resources, there are also a number of freely available services that offer single points of access to grey literature in a given field through portals and integrated search interfaces. These have a twofold relevance in collection building: a) they can act as a selection tool; and b) they can provide a relatively stable option for end user access. In this instance someone else has done the hard work of collection building for you and has dealt with the problems of selection, metadata creation and links maintenance, unless you are involved in building this type of facility yourself! For the majority of librarians the *only* task involved is locating and identifying these resources, assessing their relevance and then providing links to them for researchers to find and use. These resources tend to be developed by experts in the field and by information professionals in specialist units, leading to the creation of high quality subject specific collections. Rather than using tools to automatically collect material from the web in the way that some gateways do, the material is hand selected by experts, following defined criteria. Such services therefore provide access to authoritative information. An example of such a gateway is Science.gov[13]. The advantages of resources such as Science.gov are that they bring together material from very diverse organisations and provide search mechanisms to retrieve documents and information held in a distributed network of repositories.

4.3 Conclusion

How collections are built and what they contain will continue to evolve and develop and collecting practices will adapt to cope with changes in how information is published and disseminated. Ultimately, though, the approach to collection building in relation to grey literature chosen by individual information professionals will be dependent on the goals of the organisation. It will balance the staff and monetary resources available, the relative availability and accessibility of grey material in the subject, the aims of the organisation and the needs of the users. The aim has to be to provide coherent and integrated access to content regardless of where the content is produced and held.

We have outlined some of the current options and activities for collection building with reference to technical and government report literature in the sciences and social sciences, considering some of the benefits and problems associated with the different approaches used. In conclusion, when building collections including grey literature:

- Understand your users and have a clear vision of why you are building the collection in order to provide the right resources in the right format for your users.

13 Science.gov available at : http://www.science.gov/

- Understand the current transition from ownership of locally stored resources to the provision of access to resources, with the 'collection' offering a relevant mix of print and electronic owned, subscribed to and linked to content.
- Be pragmatic about the acquisition of material and pick approaches that suit your users and organisational needs. In reality all collection building will comprise a mixture of approaches; there is no 'one size fits all' when working with grey literature.
- Produce collection development policies, statements and selection guides to make explicit the rationale for the inclusion of content and to explain the exclusion of certain resources. Include criteria used for decisions on whether to own, subscribe to, or link to resources.
- Understand the legal environment that you are working in and what restrictions there may be on access, storage and re-use of both print and electronic content. Working with grey literature available free on the Internet is more complicated than operating in a purely print or purely commercial environment.

Grey literature has always been difficult to collect and manage and will continue to be so. However, the challenges are not unique to grey material and similar issues are being faced by librarians dealing with licensing access to mainstream commercially produced online resources. Perhaps the differences between grey and non-grey material are less clear than they once were.

Collection building in a grey digital world opens up opportunities for librarians and information professionals to develop new skills, provides fertile soil for growing partnerships and collaborations, and most importantly opens the door for the creation of rich and varied collections of resources and development of enhanced access to content for users and researchers. There is still a role for libraries in the building of collections and the inclusion of grey literature in them. What is changing is the traditional notion of what a library collection looks like and the mechanisms for creating it.

References

Lehman, H. and Webster, J. Describing Grey Literature Again: A Survey of Collection Polices. Publishing Research Quarterly, Spring 2005, pp. 64-72

Lawrence, S. et al. Persistence of Web References in Scientific Research. Computer, 34(2), February 2001, Pp 26 – 31.

Snyder, H. Libraries and digitization: forced marriage, marriage of convenience or love match? Readex Digital Instituite, Chester, Vermont, 2008. Unpublished conference presentation

Kerr, L. Intute repository search project. ALISS Quarterly, vol.3, no. 4 2008, p. 17-19. – ISSN 1747-9258

Reese, T. LibraryOn: in search of a library platform. Readex Digital Instituite, Chester, Vermont, 2008. Unpublished conference presentation

Nieminen, S. Bradford Scholars: an online repository for the University of Bradford, ALISS Quarterly, vol.3, no. 4 2008, p. 20-23

Oppenheim, C. Electronic scholarly publishing and open access. Journal of Information Science, vol.34, 2008, p. 577-590.

CABI, CAB Abstracts to include full text. Media Release, 27 October 2008 http://www.cabi.org/datapage.asp?iDocID=1216

Pickton, M. Gathering NECTAR at the University of Northampton, ALISS Quarterly, vol.3, no. 4 2008, pp. 33-38.

Chapter 5

Institutional Grey Literature in the University Environment

Gretta E. Siegel, Portland State University, USA

Historically, attention to grey literature in the academic library was focused on external collections – documents produced by government agencies or research centres. Little, if any, systematic attention was paid to the grey literature that was produced on university campuses. The advent of the Web, while bringing more interest to grey literature in general did not change this situation much. However, the trend toward the creation of institutional repositories has caused a considerable shift in interest. The formalization of collecting, processing, and integrating academic institutional grey literature should be critical to the mission of the University, regardless of format, and regardless of the existence of an active institutional repository. This chapter reviews a study on academic grey literature from earlier in the decade and provides an updated perspective.

5.1 Introduction

In the academic environment, there is an extraordinary emphasis on peer-reviewed, formally published literature. This makes sense to the teaching faculty, as their careers, in a 'publish or perish' environment, depend on this publishing model. Professors are evaluated, tenured (or not), and promoted based, to a great extent, on their output of peer-reviewed publications in high impact journals. Thus, it also makes sense that they lead their students to believe that this is the only literature worthy of consideration for inclusion in research papers, and by extension, this is the primary literature that academic libraries invest energy into, when developing collections. Another reason why grey literature has mostly been treated as 'other' by academic libraries, is simply because of a lack of familiarity. In general, this is not a subject dealt with in formal library training. Excellent cases have been made for inclusion in an LIS curriculum (Gelfand, 1998; Aina, 1998), and headway has been made in this area only recently (Farace et al., 2008).

Historically, when grey literature (other than theses, dissertations, and conference proceedings) was intentionally collected, it was most likely collections of external reports, those produced by government agencies or research institutes. In

some libraries these collections were housed as stand-alone collections, whereas in others, they may have been integrated. As more and more of these reports have now been digitized, and as current ones are 'born digital', the issues around physical integration diminish, but the issues around collection, processing, and integration into a library's holdings remain. While this is a worthy discussion, the focus of this chapter is on the grey literature produced within the university itself, though much of what is presented here could be applied to the management of external collections as well.

So the question is, if the commercially published journal literature is of such prime importance to those in the academy, would people at universities be engaged in the production of grey literature, and if so, why? And if they are, does the library collect it, and if so, how? This question was investigated in a study done some years ago at Portland State University (PSU), in Portland, Oregon (USA) (Siegel, 2004).

5.2 Review of Study and Outcomes

The study encompassed two different assessments, one was an investigation of the scholarly grey literature produced on the PSU campus, and the second was an assessment of how well we were providing bibliographic access to this body of literature. The survey instrument used to gather the initial information is given in Appendice 1.

The results of the study can be briefly summarized as follows:

1. Institutional grey literature was being, and had been produced on campus for quite a long time.
2. The library holdings included an assortment of these reports, and it could be inferred by the holdings that the library had catalogued whatever had been given to them.
3. There was no coordinated effort for the collection of these reports.
4. Grey literature was being produced on campus in virtually every discipline, with most of it coming from the social sciences.
5. The majority of the grey literature was coming from Centres and Institutes on campus.

Small but significant amounts of grey literature were also emanating from academic departments.

There was no collection development policy regarding institutional grey literature, and no established protocol for acquiring or cataloguing this material. This begged the question that if university libraries are asked to collect, catalogue, and house grey literature collections that are externally produced, though of interest to the primary and secondary clientele of the library, then shouldn't they prioritize the collection of that which is produced by the home institution? Since it cannot be anticipated that some 'other' university will be interested in collecting all that is produced on one's campus, is it not important that university libraries

capture as much of this locally produced scholarly literature as possible, regardless of format?

The reluctance of some academic colleagues to embrace the importance of this was overcome, in part, by providing a clearly articulated definition and scope of exactly what types of documents would be of interest. Almost every paper on grey literature cites the widely accepted definition: "that which is produced on all levels of government, academics, business and industry in print and electronic formats, but which is not controlled by commercial publishers", yet for any project, this definition must be refined in a way that makes sense for the scope of the project. For our university, and our early foray into formally addressing grey literature, the refined operative definition became: that which is produced BY faculty or staff IN the university, FOR THE PURPOSE of sharing scholarly information with others. This definition precluded the consideration of many things produced in academia which would be more appropriate to a university archive, or which would be seen as ephemeral, or which would open up the infinite realm of student produced literature. As is true of most universities, theses, dissertations, and conference proceedings were already being systematically collected and catalogued, so they were not of special concern in this case.

As a result of this study and of advocacy on this issue, it was concluded that if material is worth collecting and worth cataloguing, then it should be as easily located as anything else held by the library. To that end, the library made some positive changes. We enhanced the roles of the subject selectors, in addition to their liaison relationships along departmental fund lines, by assigning liaison relationships to each Centre and Institute on our campus. It became part of one's collection development duties to maintain an awareness of any reports produced by these units and to collect them (whether in print or electronic) and get them into the cataloguing pipeline. To avoid uneven collecting practices, we added a page to our collection development policy manual that outlined criteria for grey literature selection and acquisition. The additional policy statement is shown in Appendice 2.

The next step was to provide for integration into the normal workflow. The initial study and reporting of results had been effective in getting 'buy-in' from both the cataloguing department and the subject selectors (liaisons), and together, we worked out a protocol for getting the materials into the acquisition and cataloguing workflow. After assigning each Centre and Institute to a subject selector, we then needed to identify a point of contact on the other end, who would keep us each appraised of any technical reports or other grey literature that was being produced. We created 'packets' of forms as something to use for making initial contact. Additional forms were posted on the library's website.

The forms allowed the contributing unit to submit materials to us – if they were print materials, they attached the form; for digital materials, the form had space for all pertinent information. The form also allowed for contribution of potentially useful metadata by the submitting unit. The instructions that went with the forms contained the collection development statement and the caveat that not everything submitted would necessarily be accepted. The form then allowed for

review and either approval or rejection by the subject selector, and space for date tracking. A generalized version of these forms and instructions are given in Appendices 3a and 3c. Appendice 3b is the protocol that was given to each of the librarians who were tasked with collecting this material.

5.3 An Evolving Environment

For those of us with longstanding interests in grey literature, the advent of the Web simply gave us a new tool for managing, disseminating, and increasing the visibility of this literature. Prior to this development, few librarians showed much interest, but then the Web created the ultimate in grey literature – millions of ephemeral websites. The late 1990's saw several massive efforts launched at cataloguing the web, both the visible and the invisible. This seemed ironic, especially because the people who wanted to embark on this ambitious task, were often the same ones who did not see any point in dealing with paper based grey literature. Eventually this contradiction, observed by many of us ("Isn't the Web just a huge pile of grey literature?"), was explicitly articulated by Pace (2004). As time passed, the overly ambitious, and really impossible task of cataloguing the entire Web was thankfully abandoned. However, whether one is dealing with digital or print formats, wherever they exist, it gets back to the necessary step of articulating definition and scope of what it is that we want in our collections, physical and virtual. This sentiment is echoed in the 2006 paper by Pavlov (2006), in which he argues that the increased presence of grey literature on the web should not keep us from being actively engaged in the traditional activities of collection, archiving, and dissemination.

As the attentions of academic librarians were increasingly engaged in dealing with ways to combat the scholarly communication (SC) 'crisis', the idea of institutional repositories (IRs) gained traction. While not a panacea, this was at least one way in which academic institutions could ensure access to the scholarly output of their own campuses. Of course the biggest barrier to populating these burgeoning repositories was a primary aspect of the SC crisis itself, the lack of ownership of copyright by the authors of the research. As more and more scholars are now negotiating for posting rights to their published research papers, it is becoming easier to populate IRs with formally published materials.

However, in looking for ways to quickly populate repositories, since an unpopulated repository would be a hard sell to scholars, IR project managers, more often than not associated with libraries, developed a sudden interest in institutional grey literature. While the concurrent education of faculty regarding authors' rights was in process, we could meanwhile be collecting materials that did not have sticky copyright issues attached to them. A perfect example of this newfound interest in grey literature for the purpose of getting an IR off the ground is discussed in two related papers by Souloff et al. (2005) and Bell et al. (2005).

The Souloff paper discusses a study done with the help of the subject librarians at the University of Rochester who were found to "…have a depth of knowledge about the grey literature used in their own disciplines that is extensive, hard won, and valuable." One of the primary purposes of the study was "…to identify the departments and disciplines that are most likely to be early contributors [to the IR]." In this initiative, the authors see the IR as an important tool for "…disseminating the grey literature produced within the university by our own scholars."

The Bell article goes on to discuss the widening role of library liaisons, in this case, to help populate the repository. In the article previously cited Siegel (2004), the case was also made for this widening role for library liaisons, however, the purpose was not to populate any particular discovery tool or archive, but indeed to provide access to material that previously had little or no bibliographic access – institutional academic grey literature.

While I will not make the argument that institutional grey literature does not belong in a repository, I will make the argument that I made before the advent of IRs, which is that institutional grey literature should still be collected by university libraries and fully integrated into the library catalogue, whether or not they are also deposited in repositories. Several of the articles cited in the following discussion will, I believe, reinforce this view.

One advantage of the trend of populating IRs with grey literature is that studies, such as that done by Schöpfel & Stock (2009) can be conducted whereby analysis of different types of repository content and usage are looked at. In addition to finding that half of the open archives in France were owned or hosted by institutions of higher education, and that 67% of these higher education archives showed (by design) a strong academic interest in increasing the visibility of the institutions' scientific production – they also report for one particular archive, the IFREMER archive, while containing twice as much white material as grey, that the grey material was downloaded on average seven times more often.

What this underscores is the age-old observation, that grey literature is indeed useful for research; what it illustrates is that if access is provided, it will be used. In their conclusion, the authors observe that adequate bibliographic control, and therefore access, for grey literature in the open archives that they surveyed was lacking. So this gets back to the argument of exactly how access should be provided. With federated searching of repositories available, through such programs as OAIster or Google Scholar, one could argue that indeed repositories are the place for institutional grey literature, with the caveat that metadata standards could use some improvement.

Some of you will recall that when the Internet came along, there were those who argued that we no longer needed libraries. With IRs on the rise, one could argue that we do not need to include grey literature in our catalogues, as IRs will now be the logical home for them. Conclusions to the contrary can also be drawn. Unless and until repositories are completely integrated with our catalogues, they will stand as separate discovery tools. Repositories, other than those that are being designed more as 'collaboratories' (the minority), really serve the purpose of an

institutional archive of scholarly digital output, similar to how an article repository, such as JSTOR, preserves access, but is less useful as a primary search engine for discovery than is a comprehensive subject database.

The primary discovery tool for what a university library owns, or has access to, is the library catalogue, and it can be argued that this is the place where institutional grey literature must be catalogued and integrated. Note above the mention of scholarly digital output. Just as all commercial publications are not published digitally, neither is all grey literature. Though this argument may fall flat in the reality that MOST currently produced grey literature is indeed born digital, it would take significant effort and resources to digitize all of the existing grey literature that indeed, should be captured, collected, acquired, catalogued, etc.

In another article, Kargbo (2005) cogently argues the value of grey literature collections to the mission of the university. However, he uses that argument as a means for leveraging more funding and staffing. Rightly, he argues that the value in grey literature lies not in its usefulness as instructional tools, but in its potential for research. The article also notes "there is a bewildering profusion of technical activities associated with such materials…" I would posit that there is no need for this bewildering profusion, if we can simply adopt the attitude that this is material that needs to be catalogued and integrated just like any other material. And in doing so, the discrete argument for additional staffing and funding for dealing with a separate body of literature vaporizes. The point is made that "…there should be no barriers in dealing with this type of collection in academic libraries." And that librarians "…should be proactive in dealing with this type of literature in the respective institutions."

In the theoretical portion of the paper previously cited by Pavlov, there is discussion of the supply side of grey literature in the post-modern context. He points out the trend that by now we should all be aware of - that of the commodification of scientific information. Due to this trend, there is a lack of funding for the kinds of scientific research that historically has produced grey literature. He concludes that because of these trends, scientific grey literature in particular requires extra attention for funding of collection, archiving, and dissemination (i.e. for libraries) precisely because the anti-scientific post-modern market paradigm pushes us away from this.

So, while both of these articles argue for increased funding, the pragmatic approach would be a model that strongly considers integration, in order to reduce or remove the above-mentioned barriers. As long as we define this material as 'other' and in need of being kept as separate collections, we perpetuate this problem. While indeed, cataloguing of grey literature will involve a lot of original cataloguing, by contributing this metadata to bibliographic utilities, it will only need to be done once, and subsequent cataloguers will be pleasantly surprised to find that they only need to add holdings information. The fact that doing so may increase the general workload, and thus an increase in cost, is not lost, it simply becomes subsumed in any negotiations for adequate funding and staffing for the library, to carry out its mission. It seems that this will be more effective, especially in lean

economic times, as activities seen as 'special projects' are generally the first to be eliminated.

5.4 Some Comments on Integration

We have been in a place for awhile where library users would prefer 'one stop shopping' – all resources available through a single interface, and while good arguments can be made for having different interfaces for optimal retrieval of different types of resources, there is no doubt that we are heading in a unified interface direction. Interestingly though, we are doing this multi-directionally – enhancing catalogues with access to journal literature, more journal databases indexing books, repositories including multi-media, etc. It is clear that integration enhances the richness of any resource. What we will be left with in the end is anybody's guess. Integration across institutions and countries is also critical to developing a richer environment for comprehensive retrieval.

Dijk et al. (2008) describe a national program in the Netherlands, DAREnet, which integrates digital academic repositories across the country. It includes ALL universities, whereby all of the publicly funded research is deposited as well as all of the national scientific research organizations. This is their 'green route' to open access publishing. To further enhance the portal to Dutch scientific research, DAREnet is now being integrated into NARCIS (the National Academic Research and Collaborations Information System), which provides multi-layered information about national scientific research – thus creating a national union database which will allow for in-context searching of publications. And ultimately, this system will be linked into the DRIVER project – the Digital Repository Infrastructure Vision for European Research, a project that so far has eleven European countries on board.

The DRIVER project is described further in a paper by Vernooy-Gerritsen et al. 2009). The stated aim of the DRIVER project is to create an interoperable, trusted, and long-term repository infrastructure for the European community. The article looks at this project from the perspective of three stakeholders – the authors, the institutions, and information users. As of 2008, the paper reports, nearly half of the universities in Europe have implemented an Institutional Research Repository (IRR), as defined as those 'containing research output from contemporary researchers' – a refinement in definition which sets these apart from archives and heritage collections. In an analysis of the content of the repositories, it was found that overall, 33% of the items in the IRRs were full-text records, and within this 33%, 62% of the records are grey literature (theses, proceedings, working papers, etc.).

This evidence supports the claim made earlier in this paper, that grey literature is indeed the 'low hanging fruit' for populating repositories. Also in this paper, there is a brief discussion regarding the pros and cons for the variable workflows in play for deposit. Grey literature is often referred to as 'fugitive literature'

or 'the stuff that falls through the cracks'. It seems ironic, that these widespread efforts to develop infrastructures to help capture this literature would have such disjointed workflow for collection development, thus allowing whole new ways to lose these important documents. So, though this clearly is a temporary hurdle facing this particular project, it brings to light the importance of having a well-documented workflow for the collection of institutional grey literature.

Whether or not something similar to the Portland State template is adapted for catalogue integration or for repository deposit, the point is to have a protocol for workflow that involves the assignment of metadata, some collection development vetting process, and pathways for problem resolution. At the same time, an integrated process that does not place undue demand for an increase in funding or staffing, is less likely to be a target for 'cuts' in lean economic times.

European initiatives, at least compared to those in the United States (U.S.), seem to grow from a general culture, and specifically, a scientific and academic culture of centralization. The highly integrative model that we see in the DRIVER project, and the smaller projects that feed into it, are natural outcomes of this culture, and can work exceedingly well in countries and continents where scientific research is more centralized.

In the U.S., the world of research is far more fragmented. It could still be fairly far into the future before all of the scientific research conducted in the U.S. - in the universities, national research institutes, state agencies, etc. will share a common portal for discovery. Realizing the power and feasibility of such projects though, will hopefully fuel efforts at any level and any opportunity for integration.

Currently, the most widely used bibliographic utility in the U.S. is OCLC, where the front-end union catalogue product is known as WorldCat. A trend that we are currently experiencing is the integration of academic library catalogues with WorldCat, thus giving us the 'WorldCat Local' product as our home catalogues. As we move in this direction, we begin seeing the integration that users have been asking for – that of books and journal articles that previously needed to be searched via separate portals or discovery tools.

While article coverage is far from comprehensive with this product, it does belie a trend, the direction of which is obvious. In order for an item in the local catalogue to be included in the WorldCat Local catalogue, it must however have a linking identifier, in this case, an OCLC number. OCLC numbers are assigned to items as they are catalogued into the utility. Thus, grey literature which is deposited into repositories, but NOT properly catalogued into the system, meaning for most of us, OCLC, will be lost from this opportunity for discovery.

In a project described by a group of veterinary librarians (Jaros et al., 2008), a contemporary case is made for the preservation of relevant grey literature that was NOT born digital, that is very valuable to the profession and study of veterinary medicine, and that is in danger of being lost. The article echoes the argument previously made, that there must be "vigilance in collecting and preserving the output of home colleges and institutions", in spite of any prescient knowledge as to whether the value of any given document will be transitory or permanent. This article also expresses the problems encountered when holdings are not reflected in

a union catalogue, such as OCLC, and agrees that the retrieval of that which has NOT been added to a union catalogue requires extraordinary time, effort, and vigilance that most cannot afford.

An additional observation made in the original Portland State article (Siegel, 2004), but one that bears repeating is that the establishment of policies and protocols for handling institutional grey literature puts a library in a far better position to take on additional grey literature collections that may be appropriate to the University, but that also may not be widely collected or maintained, such as community-based grey literature collections that are relevant to the mission of the university.

5.5 Summary and Conclusion

To summarize the points made in this chapter:

1. The collection of scholarly institutional grey literature in academic environments should be critical to the mission of the institution, and should be articulated in collection development policies of the library.
2. A comprehensive assessment of the grey literature being produced (both quantity and sources) at any institution is advised.
3. Protocols, procedures, and responsibilities should be delineated and integrated into established workflows and position descriptions. It is recommended that these include a vetting process, to ensure consistency with other collection development guidelines.
4. By inclusion into the mission, grey literature should not be treated as an 'appended' collection – integration is key to the maintenance of consistent treatment through variable economic times.
5. Sufficient studies have shown that when academic grey literature is made available to scholars, it is utilized, fairly heavily.
6. The increased presence of grey literature on the Web is not a reason to forego efforts of comprehensive collection, cataloguing, and dissemination.
7. To optimize discovery, interoperability should be a key factor in determining whether to 'locate' grey literature in the library catalog, an institutional repository, or both.

To paraphrase something expressed in the Vernooy-Gerritson (2009) article: Ideally, what we are all trying to move toward is a system of scholarly communication that functions cohesively and at a higher level – the level of 'infusion', borrowed from the IT management literature and defined by Cooper & Zmud (1990) as "increased organizational effectiveness…obtained by using the IT application in a more comprehensive and integrated manner to support higher level aspects of organizational work."

The more that we can leverage the technology, while at the same time paying attention to mission and solid workflow to accomplish the mission; and the more we pay attention to maximizing the benefit to ALL of the stakeholders – the more

we bring the scholarly communication system to a higher level of support for high level research. It is to this end, that so many innovations are directed toward, but putting energy ONLY into disaggregated pieces of the system will not achieve this. Our entire scholarly information infrastructure needs to move toward integration in every way possible.

References

Aina, L.O. (1998) The management of grey literature as a component of a library and information science curriculum. *Third International Conference on Grey Literature*, November 13-14, 1997, Luxembourg; Amsterdam: GreyNet, 251-8.

Bell, S., N.F. Foster, and S. Gibbons (2005) Reference librarians and the success of institutional repositories. *Reference Services Review* 33(3), 283-90.

Cooper, R.B. and R.W. Zmud (1990) Information technology implementation research: a technological diffusion approach. *Management Science* 36(2),123 -139.

Dijk, E., C. Baars, A. Hogenaar, and M. van Meel. (2008). Accessing grey literature in an integrated environment of scientific research information. The Grey Journal 4(2), 57-66.

Farace, D., J. Frantzen, J. Schopfel, and C. Stock. (2008) Grey Literature: A pilot course constructed and implemented via distance education. *The Grey Journal* 4(1), 41-45.

Gelfand, J. (1998) Teaching and exposing grey literature: what the information profession needs to know: examples from the sciences. Third *International Conference on Grey Literature,* November 13-14, 1997, Luxembourg; Amsterdam: GreyNet, 242-50.

Jaros, J., E. Carrigan, C.T. Boyd, V. Croft, D. Fagen, H.K. Moberly, G. Stephens, and G. Willard (2008) Saving the Elusive Past for the Future: A Case Study in Collaboration of Veterinary Libraries. *Journal of Agricultural & Food Information* 9(4), 329-340.

Kargbo, J. A. (2005) Grey Literature: the forgotten collection in academic libraries in Sierra Leone. *Information Development* 21(2), 146-51.

Pace, A.K. (2002) Black, white and shades of gray (literature) on the web. *Computers in Libraries* 22, 44-47.

Pavlov, L.P. (2006) The Problems of grey in the context of postmodern science and culture. *Publishing Research Quarterly* 22(1), 18-26.

Schöpfel, J. and C. Stock (2009) Grey literature in French Digital Repositories: A Survey. *Tenth International Conference on Grey Literature: Designing the Grey Grid for Information Society.* Amsterdam, 8-9 December 2008. Amsterdam: TextRelease., 39-53.

Siegel, G.E. (2004) Capturing Academic Grey Literature – Starting at Home, *Publishing Research Quarterly* 20(1), 62-69.

Sulouff, P., S. Bell, J. Briden, S. Frontz, and A. Marshall. (2005) Learning about grey literature by interviewing subject librarians: A study at the University of Rochester. *College & Research Libraries News* 66(7), 510-15

Vernooy-Gerritsen, M., G. Pronk, and M. van der Graaf (2009) Three Perspectives on the Evolving Infrastructure of Institutional Research Repositories in Europe. *Ariadne* 59, <http://www.ariadne.ac.uk/issue59/vernooy-gerritsen-et-al/>, accessed June 10, 2009.

Appendice 1

Survey Instrument

[Date]

Library Survey for Scholarly Grey Literature

We at the _____ library are interested in publications produced by your department, program, school, center, or institute. We are seeking scholarly or technical reports produced by regular faculty or staff, which are published here at ____ and intended for limited distribution. This would include conference papers that have been published in full proceedings of meetings, but which the library may not have acquired.

If time and funds permit, we would like to collect this material and add it to the library collection so that it will be available to students and researchers. Please note that we ar NOT interested in materials of an ephemeral nature (e.g. brochures, newsletters, administrative notes or memos, etc.), or in materials written by students or interns.

We would appreciate it if you could take a few minutes to fill out this questionnaire. Please see the reverse side for examples of appropriate items. Thank you for your assistance.

1. Name, title, and e-mail address of person completing the survey:
2. What is the name of your department, school, program, center, or institute?
3. Do you produce any reports of the type described? Yes No

If so – could you please give us the titles and authors of individual reports, or, the title of the series and an estimate of how many separate items there are within the series?

(use a separate page if necessary)

4. Do these exist in paper format, electronic, or both? paper electronic both
5. For the ones that exist in paper, would you be willing to donate 1 copy of each report to the library? Yes No
6. For the ones that are electronic, would you be interested in working out an arrangement with the library to create access to them? Yes No
7. Please list a contact person willing to coordinate obtaining these reports from your department:
8. Any comments you would like to share with us?

Thank you very much for taking your time to help us with this project. Please return to your library liaison or to [project coordinator's name, contact info and deadline date].

Appendice 2

Collection Development Policy Statement

V. Institutional Scholarly Grey Literature: It is within the mission of the library to capture, preserve, and make available the scholarly output of the institution. To this end, the library will attempt to acquire technical reports and other scholarly publications produced by PSU Departments, Programs, Centers, and Institutes. These materials will be cataloged and added to the collection, whether in print, electronic, or both. Criteria for selection is as follows:

Authorship: The primary author(s) should be PSU faculty or staff

Content: The content should be such that a person doing scholarly research might choose to cite the work

Publication: The item would generally not be published commercially, but would have been produced in a quantity intended for limited external distribution.

Examples: Technical reports, reports of studies, conference papers that have been published in full proceedings of meetings, but which the library may not have acquired.

Examples of what NOT to collect: Materials of an ephemeral nature (e.g. brochures, newsletters, administrative notes or memos, workshop notes, course schedules, etc.); materials written by students or interns.

Appendice 3a

GREY LITERATURE SUBMISSION FORM
(top section to be completed by person submitting document to library)

Title of Document	
Subject keywords (optional)	
Author Is the author PSU staff or faculty? Yes ___ No ____	
Publishing body (e.g. Department, Center, etc.)	
Document Date	
Number of pages or URL if electronic (if submitting in both forms, please provide both)	
Person to contact if we have questions (name, phone and/or e-mail required)	
Is this document published somewhere else? If so, where?	

Next Section for Library Use Only

Meets collection development criteria?	Yes No
Selector Approval (initials and date)	
Rec'd in acquisitions (initial and date, if applicable)	
Rec'd in cataloging (initials and date)	

For additional forms, go to: [*give url for forms to be downloaded]*

Appendice 3b

NOTES FOR LIBRARIANS

Protocol / Procedure for acquiring [institution name] produced scholarly grey literature for the library

Selectors will be supplied with 'starter packets of forms to be given to their department, centre, institute, etc. liaisons. The web address for getting more forms will also be given.

The person submitting the document to the library will fill out the top part of the form and will submit the form and paper document (if any) to their subject librarian.

The subject librarian will review the document in the context of the collection development policy statement (see below) and will either accept or reject the submission.

If rejected, the librarian will return the form to the unit /person that submitted it with an explanation.

If accepted in a physical format, the librarian will initial and date the form and send both the form and the document on to Acquisitions, who will create a record and then forward it to Cataloguing.

If accepted in web format only, the librarian will initial and date the form and forward the form directly to Cataloguing.

The Cataloguing department will continue past practices of classifying the document according to subject and will catalogue the document as they would anything else. The information provided on the form is meant to be helpful but not prescriptive.

Information seen as useful to possible future problem resolution will be transferred from the form to an internal note in the item record.

The Cataloguing department will retain a file of the completed forms for 2 years, at which point the retention issue will be re-evaluated.

Collection Development Policy on Institutional Scholarly Grey Literature (adopted [date]):
It is within the mission of the library to capture, preserve, and make available the scholarly output of the institution. To this end, the library will attempt to acquire technical reports and other scholarly publications produced by [institution name] Departments, Programs, Centers, and Institutes. These materials will be cataloged and added to the collection, whether in print, electronic, or both. Criteria for selection is as follows:

1. Authorship: The primary author(s) should be [institution name] faculty or staff.
2. Content: The content should be such that a person doing scholarly research might choose to cite the work.
3. Publication: The item would generally not be published commercially, but would have been produced in a quantity intended to limited external distribution.

Examples of what to collect are technical reports, reports of studies, conference papers that have been published in full proceedings of meetings, but which the library may not have acquired.

Examples of what NOT to collect are materials of an ephemeral nature (e.g. brochures, newsletters, administrative notes or memos, workshop notes, course schedules, etc.); materials written by students or interns.

Appendice 3c

Notes for Units Submitting Documents to the Library

Thank you for helping us to collect this valuable material. The policy under which we add materials (other than traditional books, journals, conference proceedings, etc.) is as follows:

Collection Development Policy on Institutional Scholarly Grey Literature (adopted [date]):

It is within the mission of the library to capture, preserve, and make available the scholarly output of the institution. To this end, the library will attempt to acquire technical reports and other scholarly publications produced by [institution name] Departments, Programs, Centers, and Institutes. These materials will be cataloged and added to the collection, whether in print, electronic, or both. Criteria for selection is as follows:

1. Authorship: The primary author(s) should be [institution name] faculty or staff.
2. Content: The content should be such that a person doing scholarly research might choose to cite the work.

3. Publication: The item would generally not be published commercially, but would have been produced in a quantity intended for limited external distribution.

Examples of what to collect are technical reports, reports of studies, conference papers that have been published in full proceedings of meetings, but which the library may not have acquired.

Examples of what NOT to collect are materials of an ephemeral nature (e.g. brochures, newsletters, administrative notes or memos, workshop notes, course schedules, etc.); materials written by students or interns.

Please use the forms that you have been given (more available from the library website) to accompany your submission. Please submit the form and if applicable, the paper document to your library liaison. The document will be reviewed by your subject librarian, who will either accept or reject the item. If you do not receive the form back, then you can assume that the item has been accepted. We will keep the form on file and soon you will see an entry in our catalog to the document. Thanks again. If you have any questions about this program or process, feel free to contact your subject librarian or_____, Grey Literature Coordinator [contact info given here].

Chapter 6

Copyright Concerns Confronting Grey Literature

Tomas A. Lipinski, University of Wisconsin, USA

This Chapter reviews legislative and case developments in the area of copyright law affecting the collection, preservation, including digitization and dissemination, of grey literature. Alternative frameworks for crafting a legislative solution to the impediments copyright law presents to these uses are discussed. Recent threats to the availability of government-generated public domain content are assessed in light of the impact on grey literatures derived from similar public sources. Finally, recent case law supporting the archiving of various online sub-literatures is reviewed, such as the disputes over caching and archiving by Google and the TurnItIn plagiarism combating service. Short of a legislative solution, the procedural elements affecting copyright enforcement are assessed to determine the legal risk in use of grey literature.

Curren law and developments are analyzed and critiqued, with assessment towards solving the copyright issues related to the preservation and use of various grey literatures. Policy failures as well as successes in the United States can assist policy makers in other countries that are part of the community of copyright nations when contemplating issues related to preservation and use of grey literature.

6.1 Information Policy Related to Copyright in a "Grey" Context

This chapter proceeds on the assumption grey literature refers to "any documentary material that is not commercially published and is typically composed of technical reports, working papers, business documents, and conference proceedings"[1] or the "quasi-printed reports, unpublished but circulated papers, unpublished proceedings of conferences, printed programs from conferences, and the other non-unique material which seems to constitute the bulk of our modern

1 Brian Matthews, Gray Literature: Resources for Locating Unpublished Research, C&CRL NEWS, March 2004.

manuscript collections."[2] In the educational context it could also include recorded lectures and other course content, student papers, thesis' repositories, etc. The dominant theme of these conceptualizations is the unpublished nature of the literature, but is this true in every case? A later section of this paper explores the issue of publication status and asks whether in the eyes of the U.S. copyright law these works are indeed unpublished, with the impact of that publication status on use and legal risk discussed. Issues related to the institutional collection and dissemination of grey and other literatures protected by copyright is of increasing interest in the United States, the European Union[3] and world-wide.[4]

There are two options pursued in the United States when crafting legislative or regulatory "solutions" to impediments that the copyright poses to the reproduction (collection, preservation, etc.), public distribution (circulation) and perhaps public display or transmission and performance (dissemination online)—the exclusive rights of the copyright owner—of protected content. The first is to offer an exemption (or more precisely an affirmative defense) for what would otherwise be an infringing use. Exemptions come in two forms; general (those available to all works in all circumstances, such as fair use under section 107[5]) and specific (limited to the particulars of the circumstance. The statute or regulation may limit the exemption by type of work, sort of use (which exclusive rights of the copyright owner are impacted), and type of user. An example is the exemption granted to libraries and archives for reproduction and distribution of certain works under section 108.

The second option is to offer users some sort of safe harbor or protection from the impact of such infringement. This is typically crafted as a limitation on monetary[6] and in some cases injunctive remedies[7] available to copyright owners. In rare

2 PETER HIRTLE, BROADSIDES VS. GREY LITERATURE (1991), as quoted in MOYA K. MASON, GREY LITERATURE: ITS HISTORY, DEFINITION, ACQUISITION, AND CATALOGUING, available: http://www.moyak.com/researcher/resume/papers/var7mkmkw.html.

3 See, GREEN PAPER, COPYRIGHT IN THE KNOWLEDGE ECONOMY, COM(2008) 466/3, available at http://ec.europa.eu/internal_market/copyright/docs/copyright-infso/greenpaper_en.pdf, i2010: Digital Libraries, High Level Expert Group—Copyright Subgroup, FINAL REPORT ON DIGITAL PRESERVATION, ORPHAN WORKS AND OUT-OF PRINT WORKS (June 4, 2008), available at http://ec.europa.eu/information_society/activities/digital_libraries/doc/hleg/reports/copyright/copyright_subgroup_final_report_26508-clean171.pdf. See also, Annex: Model agreement for a licence [sic] on digitisation [sic] of out of work prints, available at http://ec.europa.eu/information_society/newsroom/cf/itemdetail.cfm?item_id=3366 (April 18, 2007).

4 See, e.g. STANDING COMMITTEE ON COPYRIGHT AND RELATED RIGHTS, STUDY ON COPYRIGHT LIMITATIONS AND EXCEPTIONS FOR LIBRARIES AND ARCHIVES (Seventeenth Session, Geneva, November 3 to 7, 2008) (Prepared by Kenneth Crews), available at http://www.wipo.int/meetings/en/doc_details.jsp?doc_id=109192.

5 Unless otherwise noted, all section references in the test are to Title 17 of the United States Code, the codified copyright law.

6 See, e.g. 17 U.S.C. § 504(c)(2).

7 See, e.g., 17 U.S.C. § 512(j).

instance immunity from liability may be granted.[8] This chapter assesses whether the existing and emerging legal climate is amenable to the use of grey literature in the ways that libraries, archives, and other institutional organizations might desire to obtain and make accessible grey literature, through archiving, digitization, etc. The chapter explores the current and potential interplay of the two policy options in light of proposals for reform, recent case developments and also the dynamics of copyright litigation. Finally, new threats to the continued availability of some grey literatures from copyright restoration (unique to U.S. environment) and other attempts to decrease the "size" of the public domain, including licensing, are evaluated.

6.2 Section 108 Library and Archive Reproduction and Distribution

Other than fair use (discussed below) section 108 of the United States copyright law offers qualifying institutions specific reproduction and distribution rights that may be useful in obtaining and distributing collections of grey literature. Section 108 allows for the reproduction and public distribution (circulation for example) of copies or phonorecords[9] of the collection of a qualifying library and archive for preservation and security of unpublished materials or of published materials in cases of damage, deterioration, loss, or theft, or if the existing format in which the work is stored has become obsolete.

In cases of preservation and security under section 108(b), the copy or phonorecord (or copies or phonorecords, as up to three copies or phonorecords may be

8 See, e.g., 17 U.S.C. § 108(f). Immunity is also available to state entities under the Eleventh Amendment of the U.S. Constitution. *Florida Prepaid Postsecondary Education Expense Board v. College Savings Bank*, 527 U.S. 627 (1999); *College Savings Bank v. Florida Prepaid Postsecondary Education Expense Board*, 527 U.S. 666 (1999) (states cannot be sued in federal court for patent or trademark infringement). *Rodriguez vs. Texas Commission on the Arts*, 871 F.3d 552 (5th Cir. 2000) (11th Amendment immunity extends to claims of copyright infringement). However, litigation for injunctive relief is still possible. See, e.g., Moreover, the immunity would not extend necessarily to employees of the state entity. *Cambridge University Press v. Patton*, 1:2008 cv01425 (N.D. Ga., filed April 15, 2008), available at http://www.publishers.org/main/PressCenter/documents/GSUlawsuitcomplaint.pdf. See also, Kenneth D. Crews and Georgia K. Harper, The Immunity Dilemma: Are State Colleges and Universities Still Liable for Copyright Infringement? 50 JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE 1350, 1351 (1999) (discussing the impact of "immunity" upon jurisdictional, damage, other legal liability and ethical issues).

9 17 U.S.C. § 101 defines a phonorecord as "material objects in which sounds, other than those accompanying a motion picture or other audiovisual work, are fixed by any method now known or later developed, and from which the sounds can be perceived, reproduced, or otherwise communicated, either directly or with the aid of a machine or device. The term 'phonorecords' includes the material object in which the sounds are first fixed."

made) must be from a work in the current collections of the library or archive.[10] If a copy or phonorecord is made in a digital format it must not be made available to the public in that format outside the premises of the library or archives. Remote access to the material via the library or archive website is not allowed. A copy made under subsection (b) for deposit in another library or archive may be transferred to that library or archive in digital format but the receiving library or archive must not distribute the material in that.[11] This would allow a qualifying library or archive with a collection of unpublished grey report or proceeding literature of the ABC Association or the XYZ Corporation to make a copy of the collection for preservation or security purposes or even to make a complete copy of the collection for another qualifying library or archive. The library or archive could digitize these collections as well in order to increase searching capabilities of users (staff or patrons) in accessing the content. However, the digital copies may not be made available outside the premises of the library or archive, but relegated to in-house use alone.[12]

In cases of damage, deterioration, loss, or theft, or if the existing format in which the work is stored has become obsolete under section 108(c), the copy or copies made (up to three copies may be made) are subject to the same limitation on digital distribution, i.e., remote access to the material is not allowed, and the library or archive must first make a reasonable effort to obtain an unused replacement of the published work at a fair price,[13] a so-called market check. A "reasonable effort" "will vary according to the circumstances of a particular situation. It will always require recourse to commonly-known trade sources in the United States, and in the normal situation also to the publisher or other copyright owner (if such owner can be located at the address listed in the copyright registration), or an authorized reproducing service."[14] Subsection (c) applies to published works. Less allowance is offered for published works under the statute as it is more likely for a replacement to be available. As a result, a qualifying work must be in some state of decreased availability, e.g., damage, deterioration, loss, or stolen, or if the existing format in which the work is stored has become obsolete. However, once recourse to the market place has failed, reproduction and distribution may occur, but again subject to the same space limitations for distribution of digital formats, i.e., in-house use alone.[15] Why this significant limitation on web-based access to grey literature or other collections?

---

10 17 U.S.C. § 108(b)(1).

11 17 U.S.C. § 108(b)(2): "any such copy or phonorecord that is reproduced in digital format is **not otherwise distributed in that format.**" (Emphasis added.)

12 17 U.S.C. § 108(b)(2): "any such copy or phonorecord that is reproduced in digital format is …not made available to the public in that format outside the premises of the library or archives."

13 17 U.S.C. § 108(c)(1).

14 H. Rpt. No. 94-1476, 94th Cong. 2d Sess. 75-76 (1976), reprinted in 5 United States Code Congressional and Administrative News 5659, 5689 (1976).

15 17 U.S.C. § 108(c)(2).

The legislative history of the digital copying provision of section 108, added by the Digital Millennium Copyright Act,[16] indicates that Congress was concerned with infringement vis-à-vis the proliferation of digital libraries: "Although online interactive digital networks have since given birth to online digital 'libraries' and 'archives' that exist only in the virtual (rather than physical) sense on Web sites, bulletin boards and home pages across the Internet, it is not the Committee's intent that section 108 as revised apply to such collections of information…The extension of the application of Section 108 to all such sites is tantamount to creating an exception to the exclusive rights of copyright holders that would permit any person who has an online Web site, bulletin boards, or a home page to freely reproduce and distribute copyrighted works. Such an exemption would swallow the general rule and severely impair the copyright owner's right and ability to commercially exploit their copyrighted works."[17] Thus, an on-premises library or archive use of a section 108(b) or (c) digital copy is the rule.

6.3 Solving the Problem of Orphan Works

It may be that archiving and digitization, i.e., reproduction and public distribution of a work of grey literature in its entirety may be impeded by concerns of copyright infringement because, depending on the circumstances, use of the work in its entirety may be beyond fair use or otherwise not authorized by another section of the copyright law.

6.3.1 The Problem and the Public Interest

It may be that the institutional collectors of grey literature like other users of copyrighted content would be willing to contact the owner and secure permission to use the work, even if compensation of the owner is involved. However, the owner cannot be identified or located. Given the nature of the provenance of grey literature such content may be particularly susceptible to the problem of orphan works. An "orphan work" is "a term used to describe the situation where the owner of a copyrighted work cannot be identified and located by someone who wishes to make use of the work in a manner that requires permission of the copyright owner."[18] Who is the owner of reports or position papers issued by a professional or trade association, learned or scientific society? Where such reports were com-

16 Pub. L. No. 105-304, Title IV, sec. 404, 112 Stat. 2860, 2889-2890 (1998) (codified at 17 U.S.C. § 108).

17 S. Rpt. No. 105-190, 105th Cong. 2d Sess. 63 (1998).

18 U.S. COPYRIGHT OFFICE, REPORT ON ORPHAN WORKS 15 (2006).

posed by employees of the association or society the "work made for hire"[19] rules an employer would be the copyright owner.[20] However, if an outside consultant was contracted to compose the report, the consultant would as an independent contractor, retain the copyright unless the formalities of transfer of copyright to the association were followed and executed. If a learned society published the reports as part of conference proceedings then individual contributors may have retained their copyrightable interest but the society might possess a collective copyright in the annual proceedings.[21] Lines of ownership can become easily confused. Users that desire to make use of these works but under circumstances of no legal risk will forego use in the fear that the owner could one day surface and sue for copyright infringement. As copyright law is a law of strict liability, these good faith attempts do not impact liability though general efforts of good faith may impact damages.[22] "*Such an outcome is not in the public interest*, particularly where the copyright owner is not locatable because he no longer exists or otherwise does not care to restrain the use of his work."[23]

6.3.2 The Solution: Damage Remission but not Immunity (nor Exemption)

While 2913, the Shawn Bentley Orphan Works Act of 2008[24] passed in the Senate, and though engrossed in the House on September 27, the bill failed to pass in the final days of the 110th Congress. The bill would create new section 514 of the copyright law (title 17 of the United States Code). It is likely to be re-introduced

19 17 U.S.C. § 101 indicates that a work made for hire" is either "a work prepared by an employee within the scope of his or her employment" or by designation if the work is "specially ordered or commissioned for use as a contribution to a collective work, as a part of a motion picture or other audiovisual work, as a translation, as a supplementary work, as a compilation, as an instructional text, as a test, as answer material for a test, or as an atlas, if the parties expressly agree in a written instrument signed by them that the work shall be considered a work made for hire."

20 Under 17 U.S.C. § 201(b): "In the case of a work made for hire, the employer or other person for whom the work was prepared is considered the author for purposes of this title, and, unless the parties have expressly agreed otherwise in a written instrument signed by them, owns all of the rights comprised in the copyright."

21 A "collective work" is a "is a work, such as a periodical issue, anthology, or encyclopedia, in which a number of contributions, constituting separate and independent works in themselves, are assembled into a collective whole." 17 U.S.C. § 101.

22 See, e.g., *Lowry's Reports, Inc. v. Legg Mason, Inc.*, 271 F. Supp. 2d 737, 746 (D. Md. 2003) ("The fact that Legg Mason's employees infringed Lowry's copyrights in contravention of policy or order bears not on Legg Mason's liability, but rather on the *amount of statutory and punitive damages and the award of attorneys' fees*." (emphasis added).)

23 U.S. COPYRIGHT OFFICE, REPORT ON ORPHAN WORKS 15 (2006) (emphasis added)

24 S. 2913, 110th Congress, 2d Session (April 24, 2008) (Shawn Bentley Orphan Works Act of 2008.

during the 111th Congress and pass in some form. Proposed section 514 is an example of the second form of policy "solution" to a copyright "problem" as the proposed legislation addresses the problem not by creating an exemption but in limiting the so-called bottom line or damages the user-defendant faces should the owner-plaintiff surface at some later date, decide to pursue litigation, and is successful in that litigation. If the user meets the "safe harbor" requirements of the provision then the only monetary relief the plaintiff can claim is for reasonable compensation for the infringing use made of the work. Damages (actual or statutory including damage enhancement for willful violations) as well as costs and attorney fees are not available. In some circumstances no monetary relief whatsoever is available. In the instance of derivative uses injunctive relief is also limited to an order requiring attribution for continued use and reasonable compensation for past and future uses. The derivative use cannot be suspended by the court. The question is whether or not limiting monetary liability to reasonable compensation is still too much for some would-be users to afford, i.e., this user would nonetheless in spite of the possible limitation of damages still forego use of the orphan work. Thus the impact of this "solution" would not be in the "public interest," to use the language of the Report.

Reasonable compensation is defined under proposed section 514(A)(3) as "the amount on which a willing buyer and willing seller in the positions of the infringer and the owner of the infringed copyright would have agreed with respect to the infringing use of the work immediately before the infringement began." The impact should be obvious. Users of orphan works will need to obtain some evidence or documentation of what that amount might be before use of the work commences and keep that evidence or documentation should the orphan owner ever appear one day and the user need to prove qualification under the safe harbor. The user would in theory keep the evidence or documentation of the reasonable compensation for as long as the work is being used, e.g., making a public distribution of the work by having the item in the collection of the library or archive,[25] plus three years.[26]

It is also a requirement of qualification that should the owner appear the user must bargain in good faith, offering to pay reasonable compensation. So again having documentation of what this amount might be is useful especially in cases where the owner appears years after the initial infringement and there is a difference of opinion regarding what amount the owner believes is reasonable compensation. Human nature might naturally complicate this process as the owner likely was unaware (being the "parent" of an orphan work) and now discovering that

25 *Hotaling v. Church of Latter Day Saints*, 118 F.3d 199, 203 (4th Cir. 1997) (When a public library adds a work to its collection, lists the work in its index or catalog system, and makes the work available to the borrowing or browsing public, it has completed all the steps necessary for distribution to the public.").

26 The statute of limitations for copyright infringement is three years for civil actions. 17 U.S.C. § 507. Once infringing use ceases the user-defendant could still be sued for the past infringement for up to three years, i.e., until the statute tolls.

someone was infringing their work (or to carry the analogy further, the parent is reunited with their long lost child only to discover that someone has been taking advantage of them). Considering the duration of copyright in the United States there may be a lengthy period during which this information may be relevant. So for a work for which the copyright does not expire until say 2045, where the infringing use commences in 2010, lasts until 2035 when the work is deaccessioned from the library or archive collection, the user would need to keep records of what reasonable compensation would have been in 2010 for 28 years: 25 years of use plus the three years to cover the tolling of the statute of limitations.[27] Where the use is continuous, i.e., the work remains a permanent part of a library or archive collection such as making the work accessible to the public on a website, this period would be for as long as the copyright lasts or until the work is deaccessioned, plus three years!

Under proposed section 514(c)(1)(B), a nonprofit educational institution, museum, library, archives, or a public broadcasting entity (or employees of such entity acting within the scope of their employment) can reduce the monetary amount to zero if three conditions are met. First, the infringement was performed without any purpose of direct or indirect commercial advantage. This is different than a situation where the use *results* in a direct or indirect commercial advantage, only the "purpose" must be so. In other words the use could have that effect even though that was not the intent. Second, the infringement was *primarily* educational, religious, or charitable in nature. This is not the same as a "solely" standard though it must be the primary character of the use. It could be argued that this standard looks to the character of the entity as the categorizations being those employed under the federal tax code to indicate those organizations capable of acquiring not-for-profit status, however the proposed statutory phrasing suggests the nature of "the infringement" is the classification or status of the infringer. Third, after receiving a notice of claim of infringement and having an opportunity to conduct an expeditious good faith investigation of the claim by undertaking some legal assessment of its merits the infringer must promptly cease use (infringement) of the work.

The condition to cease use upon receipt of a notice of claim might dissuade some entities from undertaking digitization or making other investment associated with the use, e.g., the cost of recording-keeping as discussed above for example. If there is a possibility that a return on that investment in the work will not be realized or halted should the owner later appear and use need to cease in order to qualify for the zero compensation provision this possibility may be sufficient to dissuade potential users of orphan works. The "notice of claim of infringement" does not require that a law suit be filed rather it is akin to the notice under section 512(c)(3) that triggers an expeditious take-down or restriction of access to con-

27 For works created after the effective date of the 1976 Copyright Act (January 1, 1978), the duration of the copyright is for author's life plus 70 years, or if corporate, anonymous, pseudonymous, the lesser of 95 years from publication or 120 from creation. 17 U.S.C. § 302.

tent.[28] As required under proposed section 514(a)(1) the notice would be made in writing and include the name of the owner and title of the infringed copyright as well as sufficient information regarding the owner or their representative and the location of the infringing content.

Finally in the case of derivative works[29] or to be more precise under proposed section 514(c)(2)(B), where the infringer has "prepared or commenced preparation of a new work of authorship that recasts, transforms, adapts, or integrates the infringed work with a significant amount of original expression," the court may not enjoin the defendant's continued use. The concept of "integration" offers a somewhat broader scope of uses than contemplated by the existing statutory definition of derivative work. Moreover, the inability to enjoin continued preparation or use in essence creates a statutory license to use the work as long as the "infringer pays reasonable compensation in a reasonably timely manner after the amount of such compensation has been agreed upon with the owner of the infringed copyright or determined by the court." If the owner refuses to agree during good faith attempts at negotiation, the court may order the owner to accept the reasonable compensation and allow the use to continue. The court must also require the user to provide attribution "in a manner that is reasonable under the circumstances to the legal owner of the infringed copyright." However attribution is only required "if requested by such owner." It could be argued that inclusion of an option for court-ordered attribution for further use is superfluous as an initial condition of section 514 qualification is to provide attribution, as discussed below. It would be odd indeed for the user of an orphan work indeed to include attribution upon initial preparation of the derivative work in order to qualify for protection under section 514 but upon surfacing of the owner and failed discussions over payment for past and future use decide to no longer provide that attribution.

6.3.3 Developments in the European Union

The approach in the European Union (EU) in the words of the Commission of the European Communities is somewhat different and tending to be more receptive of what in the U.S. would be viewed as statutory or compulsory licensing, viewing the "issue of orphan works [a]s mainly a rights clearance issue."[30] Admitting a similar policy problem, i.e., how to construct a solution that includes successful

28 17 U.S.C. § 512(c)(1)(A) ("upon notification of claimed infringement as described in paragraph (3), responds expeditiously to remove, or disable access to, the material that is claimed to be infringing or to be the subject of infringing activity").

29 17 U.S.C. § 101 defines a derivative work as "a work based upon one or more preexisting works, such as a translation, musical arrangement, dramatization, fictionalization, motion picture version, sound recording, art reproduction, abridgment, condensation, or any other form in which a work may be recast, transformed, or adapted."

30 Commission of the European Communities, GREEN PAPER, COPYRIGHT IN THE KNOWLEDGE ECONOMY, COM(2008) 466/3, p. 10, available at http://ec.europa.eu/internal_market/copyright/docs/copyright-infso/greenpaper_en.pdf.

incentives to use the orphan work while recognizing the copyright owner's interests, the Commission observed: "Apart from liability concerns, the cost and time needed to locate or identify the rightsholders, especially in the case of works of multiple authorship, can prove to be too great to justify the effort."[31] The Green Paper reviewed a number of problem areas relating to copyright and digitization, including exceptions for libraries and archives and classroom/teaching. The Green Paper proposes a series of questions regarding possible future directive on the orphan works problem and solution, such amendment of the existing Directive on Copyright[32] or other harmonization of cross-border use of orphan works.

Also issued in 2008, the i2010: Digital Libraries, High Level Expert Group—Copyright Subgroup, Final Report on Digital Preservation, Orphan Works and Out-Of Print Works[33] attempts to offer general principles of the diligent search but cautions that "regulatory initiative should refrain from prescribing minimum search steps or information sources to be consulted, due to rapidly changing information sources and search techniques."[34] The Final Report indicates that any legislative "solution" to the orphan works problem should be crafted to apply to all categories of works, but in effect different sectors may need different guidelines or best practices,[35] require a thorough search in good faith, and be flexible.[36] The i2010: Digital Libraries, High Level Expert Group—Copyright Subgroup recommends that any solution, in addition to diligent search criteria, include databases ("a registry of metadata rather than a works database"[37]) and increased use of rights clearance centers which would include "licensing conditions of the work if it remains orphan following a diligent search for the rightholder." This is far

31 Commission of the European Communities, GREEN PAPER, COPYRIGHT IN THE KNOWLEDGE ECONOMY, COM(2008) 466/3, p. 10, available at http://ec.europa.eu/internal_market/copyright/docs/copyright-infso/greenpaper_en.pdf.

32 Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonization of certain aspects of copyright and related rights in the information society, *Official Journal* L 167 June, 22nd 2001, p. 10-19, (EU Copyright Directive).

33 Dated June 4, 2008, and available at http://ec.europa.eu/information_society/activities/digital_libraries/doc/hleg/reports/copyright/copyright_subgroup_final_report_26508-clean171. pdf.

34 i2010: Digital Libraries, High Level Expert Group—Copyright Subgroup, FINAL REPORT ON DIGITAL PRESERVATION, ORPHAN WORKS AND OUT-OF PRINT WORKS 15 (June 4, 2008), available at http://ec.europa.eu/information_society/activities/digital_libraries/doc/hleg/reports/copyright/copyright_subgroup_final_report_26508-clean171.pdf.

35 See, The European Digital Libraries Initiative, SECTOR SPECIFIC GUIDELINESON DUE DILIGENCE CRITERIA FOR ORPHAN WORKS JOINT REPORT (2008), available at http://ec.europa.eu/information_society/activities/digital_libraries/doc/hleg/orphan/guidelines.pdf.

36 i2010: Digital Libraries, High Level Expert Group—Copyright Subgroup, FINAL REPORT ON DIGITAL PRESERVATION, ORPHAN WORKS AND OUT-OF PRINT WORKS 15 (June 4, 2008), available at http://ec.europa.eu/information_society/activities/digital_libraries/doc/hleg/reports/copyright/copyright_subgroup_final_report_26508-clean171.pdf.

37 i2010: Digital Libraries, High Level Expert Group—Copyright Subgroup, FINAL REPORT ON DIGITAL PRESERVATION, ORPHAN WORKS AND OUT-OF PRINT WORKS 26 (June 4, 2008), available at http://ec.europa.eu/information_society/activities/digital_libraries/doc/hleg/reports/copyright/copyright_subgroup_final_report_26508-clean171.pdf.

more ambitious then the U.S. approach, though the European Union solution does not appear to be an exemption either, but may make it more likely that users and owners will find each other through identification strategies such as increased use of metadata, and in the unsuccessful instance allow for use to continue but not with out cost, i.e., through licensing. There is recommendation that the cultural, non-profit establishments receive special treatment. Whether this will include allowance for use without specific monetary outlay—in addition to the general outlay of undertaking a diligent search—is unclear. Moreover, as in the U.S. it is unclear whether these or other measures will provide the "legal certainty [] so important for cultural institutions"[38] before an embrace of the orphan work is undertaken.

A proposed (Memorandum of Understanding on Diligent Search Guidelines for Orphan Works (hereinafter, MOA) indicates that a work "can only be considered orphan if the relevant criteria, including the documentation of the process, have been followed without finding the rightsholders."[39] The MOA does not offer any concrete or discrete steps, factors or criteria, but instead offers principles to guide the development of actual guidelines or best practices, e.g., tools to identify and mechanisms to facilitate use of orphan works, initiatives to prevent the problem in the future, and annual review. Under the U.S. approach where the emphasis appears on identifying the current rightsholder even if the communication is never established, such a copyright owner would still be considered found. Moreover, under the U.S. approach if the author or owner cannot be located but a literary agent is nonetheless identified[40] or the address of a publishing house's new corporate parent is known, the work is no longer orphan. 'Locate' is not the same as 'success' in contact, which may be suggested in the European concept of having found or "finding" the rightsholder. The (U.S.) Report makes clear that once an owner is locatable, the work ceases to be an orphan[41] regardless of the ultimate resolution of the situation, e.g., author refuses permission or never even responds at all. "This area touches upon some fundamental principles of copyright, namely, the right of an author or owner to say no to a particular permission request" or the right to say nothing at all "including the right to ignore permission requests."[42] In

38 i2010: Digital Libraries, High Level Expert Group—Copyright Subgroup, FINAL REPORT ON DIGITAL PRESERVATION, ORPHAN WORKS AND OUT-OF PRINT WORKS 15 (June 4, 2008), available at http://ec.europa.eu/information_society/activities/digital_libraries/doc/hleg/reports/copyright/copyright_subgroup_final_report_26508-clean171.pdf.

39 A Memorandum of Understanding on Diligent Search Guidelines for Orphan Works (2008), available at http://ec.europa.eu/information_society/activities/digital_libraries/doc/hleg/orphan/mou.pdf.

40 U.S. COPYRIGHT OFFICE, REPORT ON ORPHAN WORKS 97 (2006) ("For example, if it is clear from a reasonable search that an author has a literary agent to whom permission requests can be sent, the fact that the user cannot specifically locate the author (perhaps because the author is doing research in Antarctica) does not mean that the search could not 'locate' the author").

41 U.S. COPYRIGHT OFFICE, REPORT ON ORPHAN WORKS 97 (2006) ("[O]nce an owner is located, the orphan works provision becomes inapplicable.").

42 U.S. COPYRIGHT OFFICE, REPORT ON ORPHAN WORKS 97 (2006).

practice the U.S. and EU would be in accord: "Not included are works whose rightsholders refuse to authorize a use or who do not reply to a request for permission."[43] A lack of response to a permission request from a user that contains words to the effect that "unless you indicate intention to the contrary we assume your lack of response is agreement with the use we propose" is generally not a circumstance under which an implied license would exist or form the basis for any defensible right of use under the copyright law.[44] As the (U.S.) Report indicates owners might choose not to respond for a number of reasons: insufficient time or resources to respond, or an incredulous offer.[45] There is no requirement in either the (U.S.) Report or H.R. 2913 to contact the owner because once located the scenario ceases to qualify as one of orphan works. Likewise under the EU approach, an orphan work is that which is still under copyright protection and "can either not be identified, or located based on a diligent search on the basis of due diligence guidelines."[46] The evaluation of the search includes both a subjective (applied to good faith) and objective, i.e., reasonable in terms of the rightsholder (applied toe the search components or guidelines or best practices).

In June of 2009 the i2010: Digital Libraries, High Level Expert Group—Copyright Subgroup met and discussed the progress of the European Digital Library, Europeana, including the preparation of a final report of the High Level Expert Group on the topic of Digital Libraries: Recommendations and Challenges for the Future. This report may also address orphan works as the "[c]larificaiton

43 The European Digital Libraries Initiative, SECTOR SPECIFIC GUIDELINESON DUE DILIGENCE CRITERIA FOR ORPHAN WORKS JOINT REPORT 3 (2008), available at http://ec.europa.eu/information_society/activities/digital_libraries/doc/hleg/orphan/guidelines.pdf.

44 *Lowry's Reports, Inc. v. Legg Mason, Inc.*, 271 F. Supp. 2d 737, 750 (D. Md. 2003) ("Mr. Thayer did not request permission to make *any* copies of the issue Lowry's sent him. Nor did he request more than a single copy of a single issue. He asked only that Lowry's make good its alleged subscription agreement with Ms. Olszewski, who, he indicated, had not received her due copy. Moreover, the copy Lowry's sent him, like every copy it sent Ms. Olszewski herself, contained clear notice of copyright. Neither from this isolated telephone call, nor from the occasional provision of historical data, could Lowry's have known that Ms. Olszewski or Mr. Thayer routinely made and distributed copies of the *Reports* to every member of the research department. Therefore, no rational factfinder could conclude that Lowry's and Legg Mason had mutually assented to such a licensing arrangement.").

45 U.S. COPYRIGHT OFFICE, REPORT ON ORPHAN WORKS 97 (2006) ("an individual author might not have the resources to respond to every request; a large corporate owner might receive thousands of such requests and it would be unduly burdensome to respond to all of them; the request may be outlandish, in that it seeks to use a valuable work for no payment or in a way clearly at odds with the manner in which the owner is exploiting the work.").

46 The European Digital Libraries Initiative, SECTOR SPECIFIC GUIDELINESON DUE DILIGENCE CRITERIA FOR ORPHAN WORKS JOINT REPORT 3 (2008), available at http://ec.europa.eu/information_society/activities/digital_libraries/doc/hleg/orphan/guidelines.pdf.

and transparency in the copyright statue of a work is an essential element in the European Digital Library initiative."[47]

6.4 Threats to the Public Domain: Possible Impact on Grey Literature

It may also be true that the character of the grey literature the library, archive or other entity desires to harvest, organize and accession, migrate through digitization or other measures, disseminate, etc. is not protected by copyright in the first instance. Some jurisdictions such as the U.S. dedicate works of the federal government to be in the public domain.[48] The decision whether to protect state publications is left to each state legislature with understandable inconsistency in the execution of that choice.[49] It may also be that the content of the grey literature is not protected by copyright because the work does not meet the "originality" requirement.[50] In other words the work is a work of fact, such as a statistical report of a government agency or similar data such as tolerances, standards, etc. It may also be that the work simply does not contain any creativity (in the "eyes" of the copyright law). For example, a series of photographs of art, sculpture, etc., of works in the public domain where the photographs attempt to offer a precise representation of the public domain work would not be protected by copyright in the first instance.[51] Some photographs can of course be protected by copyright.[52]

---

47 i2010: Digital Libraries, High Level Expert Group—Copyright Subgroup, FINAL REPORT ON DIGITAL PRESERVATION, ORPHAN WORKS AND OUT-OF PRINT WORKS 10 (June 4, 2008), available at http://ec.europa.eu/information_society/activities/digital_libraries/doc/hleg/reports/copyright/copyright_subgroup_final_report_26508-clean171.pdf.

48 17 U.S.C. § 105 ("Copyright protection under this title is not available for any work of the United States Government.").

49 *Microdecisions, Inc. v. Skinner*, 889 So.2d 871, 874-875 (2004) (Florida law authorized "certain agencies to obtain copyrights" and "permitted certain categories of public records to be copyrighted," but it gave county property appraisers "no authority to assert copyright protection in the GIS maps, which are public records"); and *County of Suffolk, New York v. First American Real Estate Solutions*, 261 F.3d 179, 189 (2d Cir. 2001) (New York public record law "did not specifically address the impact on a state agency's copyright").

50 17 U.S.C. § 102(a) provides that "[c]opyright protection subsists, in accordance with this title, in original works of authorship fixed in any tangible medium of expression, now known or later developed, from which they can be perceived, reproduced, or otherwise communicated, either directly or with the aid of a machine or device."

51 See, *Bridgeman Art Library, Ltd. v. Corel Corp.*, 36 F.Supp.2d 191, 197 (S.D.N.Y. 1999): "In this case, plaintiff by its own admission has labored to create 'slavish copies' of public domain works of art. While it may be assumed that this required both skill and effort, there was no spark of originality-indeed, the point of the exercise was to reproduce the underlying works with absolute fidelity. Copyright is not available in these circumstances."

52 Courts have developed factors to use in assessing the creative (original, thus protected) elements in a photograph. See, *Mannion v. Coors Brewing Co.*, 377 F.Supp.2d 444, 450 (S.D.N.Y. 2005): Rendition ("copyright protects not *what* is depicted, but rather *how* it is

A series of recent cases involving government information related to land and real estate demonstrate the potential threat to continued access to public domain literature of this nature and has implications for the continued access of all public domain literature. The threats come from several strategies. First, in terms of government action, from attempts by public entities to protect by copyright what was heretofore in the public domain. Second from private entities that have access to the original sources data, e.g., the entity might have been outsourced to collect or maintain the data, and attempt to exert ownership in that public domain data. Third, a public or private entity may condition access to the public domain content through a restrictive agreement such as a license. The following two disputes demonstrate these three strategies.

In *Assessment Technologies of WI, LLC. v. Wiredata, Inc.*,[53] a private entity was contracted by a public agency (county level) to "create[] only an empty database, a bin that the tax assessors filled with the data. It created the compartments in the bin and the instructions for sorting the data to those compartments, but those were its only innovations and their protection by copyright law is complete."[54] A competing company desired to access the data to create a competing database of the content. The Seventh Circuit indicated that the competitor would be free to do so and the content of the both databases would be public domain material, which both entities were free to use. The court anticipated the first entities next move: "To try *by contract* or otherwise to prevent the municipalities from revealing their own data, especially when, as we have seen, the complete data are unavailable anywhere else, might constitute *copyright misuse*."[55] This is an important concept and dangerous trend. Attempts to control public domain content through license when access to that content is limited to single or unique source—in *Assessment Technologies of WI, LLC. v. Wiredata, Inc.*, the only collected source of the assessor information was in the database that the outsourced entity created—be thwarted by a from of estoppel[56] known as copyright misuse.

---

depicted" Id. at 452 (both emphasis original, footnote omitted). Example: "lighting selection, angle of the camera, lens and filter selection." Id. quoting, *SHL Imaging, Inc. v. Artisan House, Inc.*, 117 F.Supp.2d 301, 311 (S.D.N.Y. 2000). Timing (right place, right time). Example: famous *Catch of the Day* photograph of an Alaskan brown bear catching a salmon in mid-air as the fish attempted to jump up a waterfall during spawning run. Id. at 453. Creation of the Subject. Example, famous *String of Puppies* photograph of a couple holding a brood of puppies on their laps while seated on a sofa. Id. at 454. Photographers have sued to enforce their rights. See, e.g., *Leibovitz v. Paramount Pictures Corp.*, 137 F.3d 109 (2d Cir. 1998) (parody of Vanity Fair cover shot to promote Naked Gun motion picture fair use).

53 *Assessment Technologies of WI, LLC. v. Wiredata, Inc.*, 350 F.3d 640 (7th Cir. 2003).

54 *Assessment Technologies of WI, LLC. v. Wiredata, Inc.*, 350 F.3d 640, 646 (7th Cir. 2003).

55 *Assessment Technologies of WI, LLC. v. Wiredata, Inc.*, 350 F.3d 640, 646-647 (7th Cir. 2003) (emphasis added).

56 Black's Law Dictionary (8th ed. 2004), defines estoppel as a "bar that prevents one from asserting a claim or right that contradicts what one has said or done before or what has been legally established as true."

The doctrine of copyright misuse is adopted from similar principles in patent law relating to anti-trust.[57] The concept of misuse relates to circumstance where a valid intellectual property right exists, but the owner of the right attempts to use that right to leverage some other benefit.[58] The graveman of the misuse claim whether the plaintiff, against whom the defense is charged, is engaging in activity that undermines the public policy inherent in the copyright law, the Constitutional goal of promoting creative expression. This often occurs when the plaintiff is using the copyright law to leverage an advantage in another area.[59] The resulting anti-competitive advantage is deemed a misuse of the copyright. However, there must be "sufficient nexus between the alleged anti-competitive leveraging and the policy of the copyright laws."[60] In *Assessment Technologies of WI, LLC. v. Wiredata, Inc.*, this might be the valid copyright in the database structure itself but does not extend to the pubic domain content that the databases were designed and employed to collect. The doctrine operates to bar the copyright owner as plaintiff from suing for copyright infringement, i.e., the owner's right to enforce the right is suspended during the course of the misuse. The problem is that not all appellate

57 For a thorough discussion of anti-trust applied to intellectual property and licensing see, RAYMOND T. NIMMER, 2 INFORMATION LAW §§ 11.15 – 11.35 (2007).

58 See, Lateef Mtima, *Protecting and Licensing Software: Copyright and Common Law Contract Considerations*, INTELLECTUAL PROPERTY LICENSING TODAY, American Law Institute - American Bar Association Continuing Legal Education ALI-ABA Course of Study (SM049 ALI-ABA 81, 92 October 5 - 6, 2006) ("Finally, an infringing party who cannot claim the benefit of Fair Use may argue that the copyright holder should not be allowed to recover because she has misused or abused her copyright to obtain benefits not intended by the copyright law. The defense of copyright misuse bars a culpable plaintiff from prevailing on an action for the infringement of the misused copyright. The copyright law provides only specific property rights to the copyright holder, and competitors and the general public retain the right to challenge any over-reaching in connection with those rights. Thus the copyright law forbids the use of the copyright law to secure an exclusive right or limited monopoly not granted by the Copyright Office and which is contrary to public policy to grant." Internal quotations to *Lasercomb American Inc. v. Reynolds*, 911 F.2d 970, 972 and 977 omitted.).

59 *Lasercomb America, Inc. v. Reynolds*, 911 F.2d 970, 978 (4th Cir. 1990). ("So while it is true that the attempted use of a copyright to violate antitrust law probably would give rise to a misuse of copyright defense, the converse is not necessarily true-a misuse need not be a violation of antitrust law in order to comprise an equitable defense to an infringement action. The question is not whether the copyright is being used in a manner violative of antitrust law (such as whether the licensing agreement is "reasonable"), but whether the copyright is being used in a manner violative of the public policy embodied in the grant of a copyright.").

60 *MGM Studios, Inc. v. Grokster, Ltd.*, 454 F.Supp. 2d 966, 995 (C.D. Cal. 2006) ("StreamCast primarily alleges that Plaintiffs have restrained competition in the market for digital distribution of music and movies by collectively refusing to deal with StreamCast and other file-sharing services. … StreamCast's argument is unpersuasive. Concerted boycotts may violate the antitrust laws, but the existence of an antitrust violation is a separate question from the applicability of the copyright misuse defense. Even if Plaintiffs did act in concert to refuse licenses to StreamCast and restrict competition in the market for digital media distribution, that would not have extended Plaintiffs' copyrights into ideas or expressions over which they have no legal monopoly." Id. at 997.).

courts in the United States have adopted the concept of misuse into their copyright jurisprudence. The benefit of the misuse defense is not available only to those who are a party to the offending activity, in this case the licensee of the "egregious" license terms, but to non-parties or third parties as well.[61] Patrons of a library, archive, etc., i.e. the citizenry, would be a third party as far as the entity-licensor and the library-licensee are concerned. When misuse applies, it prevents or "estops"[62] the plaintiff from asserting a claim of copyright infringement for the duration of the copyright, but it does not necessarily prevent future attempts to do so. Once the misuse ceases the copyright owner is free to pursue legal remedy.

In a case involving public access to geographical information another court discussed copyright protection for, access to and license restrictions placed upon public domain content in the context of open records laws. In *County of Santa Clara v. Superior Court*,[63] access to public domain content was curtailed as a result of increased attention to national security in light of the events of the World Trade Center terrorist attacks: "The County also asserts a public safety interest in guarding against terrorist threats, based on its contention that the GIS basemap contains sensitive information that is not publicly available, such as the exact location of Hetch Hetchy reservoir components."[64] The court concluded that the public interest in disclosure outweighed this concern.[65] As a second rationale for not allowing access the county asserted a copyright interest in the "compilation of data", i.e., a database, in this case the "GIS basemap" and that it could therefore employ a restrictive agreement on the use of the content. The court rejected the notion that state law allowed the county to protect the public domain information through use of the federal copyright law: "The CPRA [California Public Records Act] contains no provisions either for copyrighting the GIS basemap or for conditioning its release on an end user or licensing agreement by the requester. The record thus must be disclosed as provided in the CPRA, without any such conditions or limitations."[66] Both the *Assessment Technologies of WI, LLC. v. Wiredata, Inc.* and *County of Santa Clara v. Superior Court* cases demonstrate both the merit of continued access to public domain nature and the increasing control that some

61 *Lasercomb America, Inc. v. Reynolds*, 911 F.2d 970, 979 (4th Cir. 1990) ("Therefore, the fact that appellants here were not parties to one of Lasercomb's standard license agreements is inapposite to their copyright misuse defense. The question is whether Lasercomb is using its copyright in a manner contrary to public policy, which question we have answered in the affirmative.").

62 Lectric Law Library provides the following comment on the concept of estoppel in law: "An estopple [sic] arises when someone has done some act which the policy of the law will not permit her to deny." Available at http://www.lectlaw.com/def/e040.htm

63 *County of Santa Clara v. Superior Court*, 89 Cal.Rptr.3d 374 (Cal. App. Dist. 6, 2009).

64 *County of Santa Clara v. Superior Court*, 89 Cal.Rptr.3d 374, 393 (Cal. App. Dist. 6, 2009).

65 *County of Santa Clara v. Superior Court*, 89 Cal.Rptr.3d 374, 395 (Cal. App. Dist. 6, 2009) (Independently weighing the competing interests in light of the trial court's factual findings, we conclude that the public interest in disclosure outweighs the public interest in nondisclosure.").

66 *County of Santa Clara v. Superior Court*, 89 Cal.Rptr.3d 374, 400 (Cal. App. Dist. 6, 2009), relying on relying on *Microdecisions, Inc. v. Skinner*, 889 So.2d 871, 876 (2004).

private and public entities attempt to assert over such content. As both public and private entities search for ways to maximize return in a challenging environment (economic, political, etc.), such sources of information will continue to offer attractive options for leverage, often at the expense of the public interest. It is hoped that courts continue to thwart these efforts.

6.5 Web Archiving and Fair Use

Several recent cases in the past two years have suggested that initiatives to engage in systematic archiving of content can be a fair use. In *Perfect 10 v. Amazon.com, Inc.*,[67] the Ninth Circuit concluded that Google's creation of its thumbnail index of images was fair use, commenting that "the significantly transformative nature of Google's search engine, particularly in light of its public benefit, outweighs Google's superseding and commercial uses of the thumbnails in this case." However, as the index allows users of the Google search engine to be led to infringing sources of the content, Google could be found contributorily liable: "Applying our test, Google could be held contributorily liable if it had knowledge that infringing Perfect 10 images were available using its search engine, could take simple measures to prevent further damage to Perfect 10's copyrighted works, and failed to take such steps."[68] A conclusion of fair use was also found in another case involving Google, this time its practice of automatically archiving websites unless the owner opted out. In *Field v. Google, Inc.*,[69] a district court again identified the social good that such preservation projects can achieve: "The fact that the owners of billions of Web pages choose to permit these links to remain is further evidence that they do *not* view Google's cache as a *substitute* for their own pages. Because Google serves *different and socially important purposes* in offering access to copyrighted works through 'Cached' links and does *not* merely *supersede* the objectives of the original creations, the Court concludes that Google's alleged copying and distribution of Field's Web pages containing copyrighted works was *transformative*."[70] Finally, the impact of the recent settlement by publishers and authors against Google also suggests that such archiving projects will continue to present legal challenge but through decision or settlement will be allowed to continue.[71] These developments lend support for similar efforts by institutions provid-

67 *Perfect 10 v. Amazon.com, Inc.*, 487 F.3d 701, *13 (9th Cir. 2007).

68 *Perfect 10 v. Amazon.com, Inc.*, 487 F.3d 701, *19 (9th Cir. 2007).

69 *Field v. Google, Inc.*, 412 F.Supp.2d 1106 (D. Nev. 2006).

70 *Field v. Google, Inc.*, 412 F.Supp.2d 1106, 1119 (D. Nev. 2006) (all emphasis added).

71 *The McGraw-Hill Cos. Inc. v. Google Inc.*, No. 05 CV 8881 (S.D.N.Y. filed Oct. 19, 2005); and *Authors Guild v. Google Inc.*, No. 05 CV 8136 (S.D.N.Y. filed Sept. 20, 2005). See, MOTION to Approve /Notice of Motion for Preliminary Settlement Approval (October 28, 2008); and STIPULATION AND ORDER FOR AMENDMENT OF PLEADINGS (October 30, 2008) available at http://news.justia.com/cases/featured/new-york/nysdce/1:2005cv08136/273913/.

ing similar social good by preservation of the cultural record. It may be that the same argument could be made in the case of preservation of grey literature when that collection is unique and does not exist elsewhere and the institutions serves as the sole source of the content. A final archive decision not involving Google also stands for the proposition that such initiatives offer a beneficial societal purpose and can likewise be a fair use.

In *A.V. v. iParadigms, Ltd.*,[72] the court observed that as in the case involving indexing and archiving of websites and web content, the "use of Plaintiffs' written works [is] highly transformative. Plaintiffs originally created and produced their works for the purpose of education and creative expression. iParadigms, through Turnitin, uses the papers for an entirely different purpose, namely, to prevent plagiarism and protect the students' written works from plagiarism… makes no use of any work's particular expressive or creative content beyond the limited use of comparison with other works… provides a substantial public benefit through the network of educational institutions using Turnitin. Thus, in this case, the first factor favors a finding of fair use."[73] As a result the use of the student-plaintiff's papers in the Turnitin databases was a fair use. In each of the case the use was deemed transformative and even though the entire work was taken in the instance of images in the Google cases or student papers in the *iParadigms* case the complete taking was necessary to accomplish the good purpose.

In 2009 the Fourth Circuit on appeal affirmed the fair use of student papers in the Turnitin database. In discussing the first fair use factor the court indicated that transforming uses need not alter the content in some way, but need simply put the content to a different and transforming purpose: "Plaintiffs also argue that iParadigms' use of their works cannot be transformative because the archiving process does not *add* anything to the work-Turnitin merely stores the work unaltered and in its entirety. This argument is clearly misguided. The use of a copyrighted work need not alter or augment the work to be transformative in nature. Rather, it can be transformative in function or purpose without altering or actually adding to the original work."[74] Recognizing the overlap and interconnection between the first and third and first and fourth factors the court concluded that the district court did not err in finding the use of the student papers to thwart plagiarism a fair use.[75] Again a transformative use is less likely to impact the market but only if the amount taken is no more than is necessary. This is in contrast to the recent case involving the Harry Potter Lexicon. While the nature of encyclopedias and reference guides such as the Lexicon is in general transformative, under the particular circumstances the publisher of The Lexicon: An Unauthorized Guide to Harry

72 *A.V. v. iParadigms, Ltd.*, 2008 WL 728389 (E.D. Va. 2008).

73 *A.V. v. iParadigms, Ltd.*, 2008 WL 728389. *6 (E.D. Va. 2008).

74 *A.V. v. iParadigms, Ltd.*, 562 F.3d 630, 639 (4th Cir. 2009).

75 *A.V. v. iParadigms, Ltd.*, 562 F.3d 630, 642-645 (4th Cir. 2009). "In sum, we conclude, viewing the evidence in the light most favorable to the plaintiffs, that iParadigms' use of the student works was 'fair use' under the Copyright Act and that iParadigms was therefore entitled to summary judgment on the copyright infringement claim" Id. at 645.

Potter Fiction and Related Material took more than was necessary to accomplish its good purpose.[76]

6.6 The Unknown Variable of Licensing

Licensing may impact access to grey literature (and other content for that matter) in two ways, one positive and the other negative. It may be that the grey literature is available through an online subscription to a database, posted on or via download from a website or in hard but digital format such as a CD-ROM or other disk. Each of these mechanisms might be subject to a license agreement. While librarians, archivists and other users are familiar with the concept of licensing and database the terms of use accompanying disk-based content is also subject to license. These agreements are often known as a shrink wrap agreement where the CD-ROM or other item is packaged in a box or some sort of container. The box or container is literally wrapped by a thin plastic covering that is shrunk tight to fit snuggly around the box or container.[77] Thus the term shrink-wrap is used. The significant characteristic is that the licensee does not see the terms until the package is opened.[78] Of course if the would-be licensee does not desire the product (or

76 *Warner Brothers Entertainment, Inc. v. RDR Books*, 575 F.Supp.2d 513 (S.D.N.Y. 2008). Regarding the books in the Harry Potter series: "Other times, however, the Lexicon disturbs the balance and takes more than is reasonably necessary to create a reference guide. In these instances, the Lexicon appears to retell parts of the storyline rather than report fictional facts and where to find them. " Id. 548. Regarding the companion books to the series the use is less transformative: "The Lexicon's use of copyrighted expression from Rowling's two companion books presents an easier determination. The Lexicon takes wholesale from these short books. Depending on the purpose, using a substantial portion of a work, or even the whole thing, may be permissible… In this case, however, the Lexicon's purpose is only slightly transformative of the companion books' original purpose. As a result, the amount and substantiality of the portion copied from the companion books weighs more heavily against a finding of fair use." Id. at 548-549.

77 Jonathan D. Robbins, Advising e Businesses, § 8-2.40. Acceptances on the Internet: Clickwrap, shrink-wrap and browse-wrap agreements (2006) (Shrinkwrap "Software is commonly packaged in a container or wrapper that advises the purchaser that the use of the software is subject to the terms of the license agreement contained inside the package. The license agreement generally explains that, if the purchaser does not wish to enter into a contract, he must return the product for a refund. Failure to return the product within a certain period constitutes assent to the license terms."). See also, Robert Lee Dickens, *Finding Common Ground in the World of Electronic Contracts: The Consistency of Legal Reasoning in Clickwrap Cases*, 11 MARQUETTE INTELLECTUAL PROPERTY LAW REVIEW 379, 381 (2007) ("The term 'clickwrap' evolved from the use of 'shrinkwrap' agreements, which are agreements wrapped in shrinkwrap cellophane within computer software packaging, and that, by their terms, become effective following the expiration of a predefined return period for the software (typically thirty days)." Footnote omitted.).

78 *Arizona Cartridge Remanufacturers Associaiton v. Lexmark International,Inc.*, 421 F.3d 981, 987, at n. 6 (9th Cir. 2005) (emphasis original) ("Another variant involves 'shrinkwrap licenses' on software, which impose restrictions that a consumer may discover only *after* opening and installing the software.").

service) under these terms the item may be returned. These were the circumstances of the license in numerous cases involving the purchase of computers from Gateway, Inc.[79] where the terms found inside the box upon arrival at the purchaser's home indicated that keeping the item beyond a certain time period or making a particular use of the item constitutes acceptance of the terms.[80] The cases suggest that unreasonable terms may be challenged on public policy or unconscionability grounds.[81] A variation on shrink-wrap is a scenario where rather than viewing the terms upon opening of the box or container, the terms appear upon installation (of software for example). This was the case in *ProCD, Inc. v. Zeidenberg*, the first decision to enforce a shrink-wrap agreement.[82] Although this might more accurately be called a click-wrap,[83] some courts and commentators reserve that phrase for online contracting. Courts Previous to *ProCD, Inc. v. Zeidenberg* had refused to enforce shrink-wrap licenses.[84] However, since *ProCD, Inc. v. Zeidenberg* many courts have ruled shrink-wrap licenses enforceable.[85]

---

79 *Hill v. Gateway, Inc*., 105 F.3d 1147 (7th Cir. 1996), cert. denied 522 U.S. 808 (1997) (shrink-wrap license within shipping box is valid when activated by close of 30 day return policy); Contra, *Klocek v. Gateway, Inc*., 104 F. Supp. 2d 1332 (D. Kan. 2000) (shrink-wrap license within shipping box activated by expiration of 5 day return policy not valid); and *Licitra v. Gateway 2000, Inc.*, 734 N.Y.S. 2d 389 (N.Y. Civ. Ct. 2001) (refused to uphold arbitration clause on notice and public policy grounds).

80 *Licitra v. Gateway 2000, Inc.*, 734 N.Y.S. 2d 389, 390-391 (N.Y. Civ. Ct. 2001) ("A contract results when the package is opened and the consumer uses the equipment for a specified period of time which is set forth in the written agreement. Courts have held that such a practice results in a binding contract between the parties.").

81 Robert W. Gomulkiewicz and Mary L. Williamson, A Brief Defense of Mass Market Software License Agreements, 22 *Rutgers Computer & Technology Law Journal* 335, 345 (1996) ("Rather than relying on their own negotiating skills or knowledge of the relevant law, most users are better served by relying on the contract doctrine of unconscionability, the contract principle that agreements should be construed against the drafter, the copyright doctrine of misuse, consumer protection laws, and the intense competition within the software market to obtain advantageous terms in acquiring software.").

82 *ProCD, Inc. v. Zeidenberg*, 86 F.3d 1447, 1449 (7th Cir. 1996) ("Shrinkwrap licenses are enforceable unless their terms are objectionable on grounds applicable to contracts in general (for example, if they violate a rule of positive law, or if they are unconscionable).").

83 See, e.g., *.i.Lan Systems, Inc. v. Netscout Services Level Corp.,* 183 F.Supp.2d 328, 329 (D.Mass.2002) ("You plunk down a pretty penny for the latest and greatest software, speed back to your computer, tear open the box, shove the CD-ROM into the computer, click on 'install' and, after scrolling past a license agreement which would take at least fifteen minutes to read, find yourself staring at the following dialog box: 'I agree.' Do you click on the box? … Is that 'clickwrap' license agreement enforceable? Yes, at least in the case described below.").

84 See, e.g., *Step-Saver Data Systems., Inc. v. Wyse Technology*, 939 F.2d 91, 102-03 (3d Cir. 1991); *Vault Corp. v. Quaid Software Ltd.*, 847 F.2d 255, 268-70 (5th Cir. 1988); *Arizona Retail Systems, Inc. v. Software Link, Inc.*, 831 F. Supp. 759, 763-66 (D. Ariz. 1993).

85 See, e.g., *Davidson & Assocs. v. Jung*, 422 F.3d 630, 638-39 (8th Cir. 2005); *Bowers v. Baystate Techs., Inc.*, 320 F.3d 1317, 1323-25 (Fed. Cir. 2003); *Meridian Project Systems, Inc. v. Hardin Construction Co.*, 426 F. Supp. 2d 1101, 1106-07 (E.D. Cal. 2006); *Information Handling Services, Inc. v. LRP Publications, Inc.*, No. Civ.A. 00-1859, 2000 WL 1468535, at 2 (E.D. Pa. Sept. 20, 2000); *Peerless Wall & Window Coverings, Inc. v. Syn-*

Less awareness may be present when using content available from a website. As transactions moved online, so did licensing where terms may appear during completion of the transaction or as the product or service is obtained such as through download. Prompts appear that allow the licensee to view the terms of the agreement and to assent to those terms by clicking "I agree" or some other prompt.[86] Thus the term click-wrap is used to describe agreements where as with shrink-wrap, the licensee is without opportunity to bargain and finds the transaction wrapped solely in the terms offered by the licensor.[87] The legal analysis used in the shrink-wrap scenarios is applied by the court in assessing click-wrap as well.[88] Various courts have also upheld click-wrap agreements as valid.[89] A key

chronics, Inc., 85 F. Supp. 2d 519, 527 (W.D. Pa. 2000); *Adobe Systems, Inc. v. One Stop Micro, Inc.*, 84 F. Supp. 2d 1086, 1090-91 (N.D. Cal. 2000); *M.A. Mortenson Co. v. Timberline Software Corp.*, 998 P.2d 305, 311-13 (Wash. 2000).

86 Jonathan D. Robbins, Advising e Businesses, § 8-2.40. (2006) (Acceptances on the Internet: Click-wrap, shrink-wrap and browse-wrap agreements) ("A click-wrap license presents the user with a message on their computer screen, requiring that the user manifest his consent to the terms of the agreement by clicking on an icon. The user cannot continue to view the website or buy the particular product unless and until the icon is clicked." Footnote omitted.). See also, Robert Lee Dickens, Finding Common Ground in the World of Electronic Contracts: The Consistency of Legal Reasoning in Clickwrap Cases, 11 *Marquette Intellectual Property Law Review* 379, 381 (2007) ("In such transactions, sellers have increasingly begun utilizing clickwrap agreements, whereby standard terms and conditions are displayed on the computer screen when the user attempts to access the seller's services. In a clickwrap agreement, the seller's terms typically pop up before a purchased software disc can be installed (CD clickwrap) or while a service is being requested on the Internet." Footnotes omitted.).

87 See, Rachel S. Conklin, Be Careful What You Click For: An Analysis of Online Contracting, 20 *Loyola Consumer Law Review* 325, 327 (2008) ("Clickwrap agreements required the user to ask some manifestation of his or ner intent to be bound by a contract after being present with that contract's terms, for instance by clicking a button labeled 'I agree' after viewing the terms."). See also, Scott J. Lochner, *A Legal Primer on Software Shrink-Wrap: Click Wrap or Click-To-Accept and Browse-Wrap License Agreements*, INTELLECTUAL PROPERTY TODAY, DECEMBER 2003 "Generally, click-wrap license agreements are either (i) online (i.e., over the internet) license agreements that are used when copies of software are marketed and delivered electronically, or (ii) license agreements that are part of the initialization process that occurs during the loading of software on a computer. These license agreements for software are referred to as 'click-wrap' or 'click-to-accept' license agreements because the initialization procedure requires the customer to click on an 'enter' or 'approved' icon in order to signify acceptance to the terms of the software license agreement."

88 Lateef Mtima, Protecting and Licensing Software: Copyright and Common Law Contract Considerations, *Intellectual Property Licensing Today*, SM049 ALI-ABA 81, 96 (October 5 - 6, 2006) (American Law Institute - American Bar Association Continuing Legal Education Program) (citations to cases omitted) ("Currently the courts remain divided on the issue of the enforceability of shrinkwrap licenses. Some courts continue to find them unenforceable… In general, however, it seems that a shrinkwrap license is more likely to be held enforceable where (i) there is evidence that the user is aware of the license, (ii) there is concrete manifestation of assent to the license terms or a reasonable period of time upon which assent will be inferred, and (iii) it contains commercially reasonable terms, particularly where consumers are involved. In contrast to shrinkwraps, the courts have had less diffi-

component in an enforceable click-wrap agreement is the availability of the terms prior to the click and a specific indication that clicking equals assent, i.e., "I agree", to those terms. If the user is required to "click" to enter it is very likely that an enforceable agreement governs the use of the content found on that website, a website that might include content in the category of grey literature.

If a party is not aware of a term or had no opportunity to become aware of the term then does not undertake any act of assent there is no contract. "As we have seen, standards for forming a contract concentrate on whether there are objective indicia (manifestations) of assent. In the typical online environment, assent to a contract entails assent to terms of a standard form set out by the site owner or product vendor. The assent issue involves whether the site user or product purchaser assented to the terms."[90] Where there is no manifestation of assent, courts will not hold the party to a term to which it did not agree. In *A.V. v. iParadigms, Ltd.*, the court commented: "the Usage Policy is not binding on Plaintiffs as an independent contract because Plaintiffs did not assent to the Usage Policy… In this case, there is no evidence that Plaintiffs assented to the terms of the Usage Policy. There is no evidence that Plaintiffs viewed or read the Usage Policy and there is no evidence that Plaintiffs ever clicked on the link or were ever directed by the Turnitin system to view the Usage Policy. There is no evidence to impute knowledge of the terms of the Usage Policy to Plaintiffs."[91] A similar result was reached in *Williams v. America Online, Inc.* where the terms could not be viewed until after the "click" and the court concluded that meaningful assent could not be

culty upholding clickwraps, primarily because these agreements typically require the user to indicate assent to the terms of the license before she can obtain or use the software. Whether a shrinkwrap or a clickwrap, however, a court could find an enforceable license overall but nonetheless make independent rulings as to the enforceability and/or commercial reasonableness of a specific standardized term.").

89 Compare, *In re RealNetworks Privacy Litigation*, 2000 U.S. Dist. LEXIS 6584, *6 (N.D. Ill. 2000) ("The user can then click on the License Agreement, listed separately as either 'RealJukeBox License Agreement' or 'RealPlayer License Agreement,' depending on the product, and easily print out either agreement from the file pull down menu."); with *Comb v. PayPal, Inc.*, 2002 U.D. Dist. LEXIS 16364 (N.D. Cal. 2002) (arbitration clause found "procedurally unconscionable": freeze funds, prohibition of consolidation, $5,000 cost of arbitration, venue unreasonable). See also, *DeJohn v. The .TV Corporation International,* 245 F.Supp.2d 913, 915-916 (N.D.Ill.2003) ("The electronic format of the contract required DeJohn to click on a box indicating hat he had read, understood, and agreed to the terms of the contract in order to accept its provisions and obtain the registration or reject the provisions and cancel the application. This type of online contract is known as a click-wrap."); *Koresko v. RealNetworks, Inc.*, 291 F.Supp.2d 1157, 1163 (E.D.Cal.2003) ("Plaintiff accepted the terms by clicking 'I agree' to the terms and conditions of the contract including the forum selection clause."); *Stomp, Inc. v. NeatO, LLC*, 61 F.Supp.2d 1074, 1081 (C.D.Cal.1999); and *Regency Photo & Video, Inc. v. American Online, Inc.,* 214 F.Supp.2d 568, 573 (E.D.Va.2002).

90 RAYMOND T. NIMMER, 2 INFORMATION LAW § 12.33 (2007).

91 *A.V. v. iParadigms, Ltd.*, 544 F.Supp.2d 473, 485 (E.D. Va. 2008), affirmed 562 F.3d 630 (4th Cir. 2009).

give to terms that could not be viewed.[92] Sounds logical, but some licensors have attempted to push the envelope of the concept of meeting of the minds.[93]

One characteristic of a so-called browse-wrap agreement is its occurrence exclusively in web-site settings. More important browse-wraps are characterized by obscurity regarding the terms of the agreement. Obscurity is present both in the appearance of the terms and in the mechanism of assent.[94] Often the terms do not appear in conjunction with the assent mechanism ("click here to see the terms" together with "click here to agree to the terms") but rather transport the licensee to some other portion of the website ("to view the terms click here") or appear only

92 *Williams v. America Online, Inc.*, 2001 WL 135825, *3 (Mass. Super. 2001) (unpublished) (AOL motion to dismiss denied) ("Cass, who has more than 20 years experience with mainframe and personal computers, owns and operates Cass, Inc., a provider of database and computer support services. In his affidavit, Cass describes in detail the AOL 5.0 installation process. He states that the alleged harm occurs before the user clicks "I agree". He describes a complicated process by which subscribers "agree" to the TOS after configuration of the computer has been altered. AOL sets the default for reviewing the TOS to "I agree." A customer who merely clicks "I agree" is instantly bound by the terms of a TOS she has never seen. The customer's only other option is to click off the default and select "Read Now." That option also fails to provide a customer with an opportunity to read the TOS. A customer who selects "Read Now" is presented with another choice between the default "OK, I agree" and "Read Now". Thus, the actual language of the TOS agreement is not presented on the computer screen unless the customer specifically requests it by twice overriding the default…Therefore, the fact that plaintiffs may have agreed to an earlier TOS or the fact that every AOL member enters into a form of TOS agreement does not persuade me that plaintiffs and other members of the class they seek to represent had notice of the forum selection clause in the new TOS before reconfiguration of their computers.").

93 Viewable terms that require the licensee to undertake efforts to determine when changes or up dates to the terms occur are also suspect. See, *Douglas v. Talk America, Inc.*, 495 F.3d 1062, 1065 (9th Cir. 2007). Facts: "Joe Douglas contracted for long distance telephone service with America Online. Talk America subsequently acquired this business from AOL and continued to provide telephone service to AOL's former customers. Talk America then added four provisions to the service contract: (1) additional service charges; (2) a class action waiver; (3) an arbitration clause; and (4) a choice-of-law provision pointing to New York law. Talk America posted the revised contract on its website but, according to Douglas, it never notified him that the contract had changed. Unaware of the new terms, Douglas continued using Talk America's services for four years." New terms are not part of the agreement: "Even if Douglas had visited the website, he would have had no reason to look at the contract posted there. Parties to a contract have no obligation to check the terms on a periodic basis to learn whether they have been changed by the other side. [footnote 1]" Id. at 1066. Footnote 1: "Nor would a party know *when* to check the website for possible changes to the contract terms without being notified that the contract has been changed and how. Douglas would have had to check the contract every day for possible changes. Without notice, an examination would be fairly cumbersome, as Douglas would have had to compare every word of the posted contract with his existing contract in order to detect whether it had changed." Id.

94 See, Rachel S. Conklin, Be Careful What You Click For: An Analysis of Online Contracting, 20 *Loyola Consumer Law Review* 325, 327 (2008) ("On the other hand, the terms of a browsewrap contract are often inconspicuous or even unavailable to a consumer online; a contract is accepted by performance as the consumer continues to navigate the website or uses a product or service found on the site.").

after the user scrolls several screens forward. Second, the assent mechanism itself is obscure. Rather than a precise pronouncement of assent ("to agree to these terms, click here") the licensor conditions assent so some other conduct such as use of website services ("by submitting a query you agree to be bound by the terms") such as submitting a price-quote or ticket availability query. It should be obvious that the validity of browse-wrap agreements is met with far more scrutiny by the courts.

This is the negative aspect of licenses. Terms of the license may restrict use of the content by limiting reproduction of the content to the specific user, curtailing further distribution of the content beyond the specific user or dictating how the content may be used, i.e., limiting the ability to make a public display or public performance of the content or prohibit making derivative use of the work. While it is beyond the scope of this chapter to discuss in detail the "law" of licensing one recent trend of significance is that website terms of use or so-called End User License Agreements (EULAs) are enforceable including those circumstances where use of the website can constitute assent to those terms.[95]

The Creative Commons license is likewise enforceable.[96] This is the positive side of licensing. Release of grey literature from its source (e.g., trade association, learned society or professional organization, etc.) or re-distribution of grey literature by a library, archive, etc. may be made pursuant to a Creative Commons or other license structure. Such license prohibits commercial or derivative use and in addition condition use by others on a similar commitment of open access, so-called serial licensing or in the terms familiar to Creative Commons users "share-alike" where the subsequent user (the second user) must also accept conditions similar to those regulating the first user. In this way the dissemination chain of grey literature is maintained by all uses in an open access environment and under similar rules. The extent to which grey literature is available subject to license is unknown but as with TPM use of licensing by content providers is increasing. Use of license terms can prohibit certain anti-access conduct from occurring. For example, grey literature could be made available to the public subject to license term that prohibits the placement of TPM on further uses of the content in addition to the "no commercial" use of the existing and familiar Creative Commons schema.

---

95 See, *Ticketmaster L.L.C. v. RMG Technologies, Inc.*, 507 F.Supp.2d 1096 (C.D. Cal. 2007) (automated extraction by brokers of ticket information from website): "Thus, by the Terms of Use, Plaintiff grants a nonexclusive license to consumers to copy pages from the website in compliance with those Terms. Inasmuch as Defendant used the website, Defendant assented to the terms." Id. at 1108.

96 See, *Jacobson v. Katzer*, 2008 WL 3395772 (Fed. Cir. 2008) ("We consider here the ability of a copyright holder to dedicate certain work to free public use and yet enforce an 'open source' copyright license to control the future distribution and modification of that work." Id. at *1. The court concluded that Creative Commons type licenses are enforceable under the copyright law ("the terms of the Artistic [Creative Commons] License are enforceable copyright conditions." Id. at *8.). See also, Lydia Pallas Loren, Building a Reliable Semi-commons of Creative Works: Enforcement of Creative Commons Licenses and Limited Abandonment of Copyright, 14 *George Mason Law Review* 271 (2007).

In light of recent case law in the United States these limitations, in this instance to the benefit of the public, would be enforceable.

6.7 Conclusion

The expanded collection and dissemination of grey literature (as well as other works protected by copyright) through archiving and digitization is bolstered by recent case law establishing the circumstances under which such initiatives can be a fair use under U.S. copyright law. In addition legislative reform is under way to increase range of use rights available to institutions regarding protected content including grey literature. Moreover, the particulars of copyright enforcement may also work to minimize the legal risk in remaining circumstances. Finally, licensing may prove to be a bane as well as a boom to the continued access, preservation and use of grey literatures as more and more content generation is moved online and providers adopt a method of distribution based on licensing models.

Part I, Section Three

Channels for Access and Distribution of Grey Literature

The U.S. Government's Interagency Gray Literature Working Group (IGLWG)[1] defined grey literature in 1995 as "foreign or domestic open source material that usually is available through specialized channels and may not enter normal channels or systems of publication, distribution, bibliographic control, or acquisition by booksellers or subscription agents".

The goal of this section is to provide insight in the distribution channels of grey literature, especially in the field of academic publishing. The focus rests on digital information and open access.

In the introductory chapter to this monograph, we stated that the proportion of grey documents in relation to commercially published documents on the Web continues to increase. This development seems closely linked to the production of grey literature in digital environments, as well as to retrospective activities commensurate to republication.

We further purport that open archives will provide more tailored services and functionality for at least some segments of grey literature namely preprints, doctoral theses, and reports. We mention these three types of grey documents, because they have come to form special collections more visible than ever in repositories.

The first chapter in this section presents an overview of production and dissemination channels for Ph.D. theses. Stock and Paillassard's work is primarily based on the situation in France, however, their study also explores several national and international projects and initiatives on electronic theses and dissertations (ETDs) in Europe and the United States. The authors ascertain that while "technical developments have greatly facilitated the dissemination of ETDs, (…) legal issues (…) became an obstacle." And, they conclude that "the growing complexity of the ETD landscape calls for explicit policies in the future to inform the user of a given repository on deposit, validation, access and reuse of a thesis."

The second chapter in this section rest assures that "there is no doubt (…) GL is at *home* in open archives". Luzi sets out a comprehensive and well-documented

1 This working group became dormant in early 2000. However, a new working group is currently in the process of being launched under the name Grey Matters USA. A leadership group was formed during the Eleventh International Conference on Grey Literature held in the Library of Congress on 14-15 December 2009.

overview on the evolution of grey literature during the past two decades moving from print to digital formats and from library holdings to open repositories. Luzi's study helps in understanding the recent history of scientific information. Coverage of the preprint culture, scientific artifacts, institutional repositories, and interconnected knowledge networks are primary features in her line of discourse. In conclusion, Luzi examines the relationship between grey and conventional literature in an open environment. She remarks, "the coexistence of GL with conventional literature actually provides an ideal, complete coverage of the research results of any given scientific institution or disciplinary community. (…) The distinction between GL and conventional literature is becoming increasingly blurred (…) the main difference vis-à-vis conventional literature is inherent mainly in the fact that GL is not subjected to any formal peer-review process."

For nearly a quarter century, a number of national libraries and research centres in Europe maintained a network for the collection and dissemination of grey literature built around the SIGLE (System for Information on Grey Literature in Europe) database. Our final chapter in this section describes the integration of the former SIGLE records into a new open access project called OpenSIGLE[2]. The authors discuss the roles of the service provider and data provider, present examples of usage statistics, and conclude with a research proposal that would explore the creation of an e-infrastructure in order to serve the OpenSIGLE Repository. "The outcome of this project would support and strengthen policy development for infrastructures in the field of grey literature, where open access to their collections and other knowledge based resources stand central."

Five years ago, Willinsky stated that open access to information is a common good[3]. And for a couple of reasons, this principle appears to apply more to grey items than to journals and books. First of all, because a significant percentage of grey items are produced by public bodies, and secondly because they are already "off-commerce" *i.e.* not controlled by commercial suppliers. For this kind of publisher, the economic challenge or risk of "going OA" seems rather minimal. And, the studies in this section seem to corroborate that public scientific information centres are already more or less involved in open access projects with grey literature.[4]

With this in mind, we urge that the readers consider the following three questions as they proceed through the chapters in this section:

What is (could be) the impact of open repositories on grey collections? Does the open access movement improve the search and retrieval of grey documents?

---

2 http://en.wikipedia.org/wiki/OpenSIGLE

3 J. Willinsky (2005). *The Access Principle: The Case for Open Access to Research and Scholarship (Digital Libraries and Electronic Publishing)*. The MIT Press.

4 See C. Boukacem-Zeghmouri & J. Schöpfel (2006). `Document supply and open access: an international survey on grey literature'. *Interlending & Document Supply* 34(3):96-104. J. Schöpfel & H. Prost (2009). `Document supply of grey literature and open access: an update'. *Interlending & Document Supply* 37(4):181-191.

And finally, how can one improve the referencing and access to grey documents deposited in open repositories?

Chapter 7

Theses and Dissertations

Christiane Stock and Pierrette Paillassard
INIST-CNRS, France

7.1 Thesis and/or dissertation – terms and scope

According to Wikipedia[1]: "A dissertation (also called thesis or disquisition) is a document that presents the author's research and findings and is submitted in support of candidature for a degree or professional qualification. In some countries/universities, the word thesis is used as part of a Bachelors or Masters course, while dissertation is normally applied to a Doctorate." However, this usage is non consistent throughout the countries.

A doctoral thesis is the result of 3-4 years of research and the first valuable document in the career of a researcher. At the same time it is an administrative document necessary to obtain the doctoral degree. In some disciplines theses are considered as the result of teamwork and appear in the list of publications of the laboratory where the research was done (Mermet et al. 1998)

Limitations of the "study": the main focus of this chapter will be set on doctoral theses or doctoral dissertations. In many countries Master theses are not considered worth an effort of dissemination in the academic context. Considerations about the quality of the content may influence the decision of non-dissemination. We also exclude initiatives outside the academic or research context, e.g. initiatives taken by students to disseminate their works through associative websites. These private initiatives may not provide stability over time nor make any control of the input.

7.2 A short history of dissemination

Where can doctoral theses or dissertations be found?

Since they are produced in the universities, they were deposited in the university library and included in the library catalogue. Copies were made available to other universities through inter-library-loan.

1 Wikipedia, http://en.wikipedia.org/wiki/Dissertations

Besides library catalogues doctoral theses were included early on in monthly/annual bulletins or indexes, then in databases on a national or international level, such as the catalogue of the British Library and its monthly bulletin of "British Reports, Theses and Translations".

In France a four-level national network for theses was established following a decree published by the French Ministry of Education in 1985. It included among others the systematic reproduction and dissemination of the documents on microfiche the dissemination of a copy to all French universities as well as the systematic referencing through a national database: Téléthèses. The latter has been integrated into the national union catalogue for university libraries, SUDOC*.

UMI*: the most important and long-lasting initiative worldwide is found in the United States: University Microfilm International (UMI*) in Ann Arbor collects (since 1938) abstracts and full texts of doctoral dissertations from North American and European universities which are indexed in Dissertation Abstracts International (paper and microform), and its internet database Dissertation Abstracts Online, now ProQuest Dissertations & Theses (PQDT) (UMI 2009). More than 90% of the doctoral dissertations accepted by North American universities are covered by UMI*. ProQuest UMI Dissertation Publishing has published over 2 million graduate works in 70 years.

SIGLE*: in Europe the SIGLE* database (System for Information on Grey Literature in Europe) played a major part in providing access to doctoral theses. Produced by EAGLE (European Association for Grey Literature Exploitation), it was entirely dedicated to the collection and dissemination of grey documents to which theses and dissertations belong. Its 15 members provided records facilitating the identification of the documents and held a copy for document delivery on demand. More than 275,000 records referenced doctoral theses, covering the period from 1980 to 2004, representing about one third of the database. The majority of the database is now in open access at http://opensigle.inist.fr.

NDLTD*: is another milestone in the doctoral theses universe. The Networked Digital Library of Theses and Dissertations (NDLTD*) was created in 1996 by Virginia Tech (Fox 1996), funded by the U.S. Department of Education and the Southeastern Universities Research Association in order to improve graduate education. The outcome of the project and further activities have greatly influenced the transition to electronic dissertations throughout the world. Work on electronic theses and dissertations started as early as 1987 when Virginia Tech developed the first SGML Document Type Definition (DTD) for theses and dissertations (Fox 1996).

Today's proclaimed aim of NDLTD* is to promote the "adoption, creation, use, dissemination and preservation of electronic analogues to the traditional paper-based theses and dissertations."[2] It federates more than 100 members and offers access to almost 800,000 online documents through the NDLTD* Union Catalog[3] based on OAI-PMH.

2 NDLTD, http://www.ndltd.org

3 NDLTD Union Catalogue, http://alcme.oclc.org/ndltd/

The NDLTD* website provides a host of tools and guides for electronic theses and dissertations developed by the federation.[4]

A comprehensive guide (ETD Guide 2009) covers on a technical level all issues for universities and students.

The ETD metadata scheme "ETD-MS: an Interoperability Metadata Standard for Electronic Theses and Dissertations" has been adopted worldwide. It adds an element "theses.degree" to the basic Dublin Core metadata scheme with the following qualifiers[5]:

thesis.degree.name
thesis.degree.level
thesis.degree.discipline
thesis.degree.grantor

Since its first meeting in 1998, NDLTD* organizes the international conference ETD (ETD Guide 2009).

7.3 From the digital library to digital workflow - a new organisation for electronic theses

Electronic theses and dissertations can be produced in different ways. The easiest way is to digitize a paper copy. Converting a text format into a PDF file comes as a second. However, changing to the electronic version is not simply a change of the support of dissemination, but needs a whole new organization.

In France, first steps in the electronic era were taken in the 1980's-90's when universities digitized their theses, often in PDF format, and put them on the internet.

Thus INRIA*'s doctoral theses were accessed from the 90's through a simple HTML webpage on its institutional website. Since 2005, students are encouraged to submit their document to the self archiving TEL-HAL*[6].

This archive is today the most comprehensive French repository. It was created by the CCSD* and MathDoc*, one of the oldest French archives.

MathDoc* and Grisemine (now IRIS*) were among the first to digitize through paper copies. Since 1997 another of the top French engineering universities, INSA Lyon* offered its students different tools to produce electronic theses: converting a text format into a PDF file.

At the same time, a common project between Canadian and French universities (Montreal, Lyon) proposes a complete editorial chain in open source software called "Cyberdocs*". It covers aspects from document model to the conversion into a fully structured XML document using the TEI Lite DTD.

---

4 NDLTD-documentation, http://www.ndltd.org/resources

5 NDLTD-standards, http://www.ndltd.org/standards/metadata/etd-ms-v1.00-rev2.html

6 TEL-HAL, http://tel.archives-ouvertes.fr/index.php?langue=en&halsid=ulmij8bipli3u8759ggfi72ko7

Following NDLTD*, the German project Dissonline* was one of the first to tackle the problem with a holistic approach: separate workgroups dealt with metadata issues, publication tools, development of the server platform OPUS, creation of a new workflow, the provision of training for students and administration staff, and last but not least help in resolving legal aspects. Universities had to revise their graduation regulation to accept electronic versions as valid documents. Authors rights and German copyright laws were other issues that need to be dealt with.

The successful outcome of the programme is to be seen in the catalogue of the German national library which includes more than 78,000 records of electronic dissertations.

France proceeded in several steps, defining first a national metadata scheme (TEF*), before working on the workflow and server issues. Valuable information can be found on the website ORI-OAI*[7]

7.4 Dissemination of ETDs today

Next to journal articles and eprints, electronic theses and dissertations (ETDs) are for various reasons the most frequent document type found in open archives.

- ETDs are a well defined and well referenced document type. Rules for deposit and citation are generally established on a national level, and international standards exist for specific information and theses metadata, contrary to other grey documents.
- ETDs are administrative documents, and students can be "obliged" to deposit their work in an archive or repository in order to obtain their diploma.

Table 1: Source: OpenDOAR* (data collected July 24th, 2009)

Continent	Repositories registered	Repositories with theses	%
Europe	686	372	54 %
North America	404	160	39 %
Australasia	77	54	70 %
Asia	167	81	48 %
South America	68	32	47 %
Africa	24	17	70 %
Total	**1,426**	**716**	**50 %**

According to the OpenDOAR* - the Directory of Open access repositories registry - 716 out of 1,426 repositories (50%) contain theses. Only for North America the

7 ORI-OAI, http://wiki.univ-paris5.fr/wiki/ORI

percentage is significantly lower. Here, we cannot distinguish between master's theses and doctoral theses.

It should be noted that not all repositories with ETDs are registered in OpenDOAR*, while some sites included in the table may still be at an experimental stage and contain either very few records or only test records. The data above don't provide any information on the real number of theses available through those repositories.

7.4.1 From auto-archiving to portals

The first archives for scientific publications including ETDs were based on the principle of self-archiving or deposit by the author, intended for fellow researchers. The documents submitted were reviewed for formal aspects or their contents; and the metadata were entirely supplied by the authors. An early example in France is TEL* (thèses en ligne), which has since been included into HAL*.

A second phase saw the creation of institutional repositories (IR) including the scientific production of a given university or research institute. IRs fullfill the role of a showcase to the world, but they are also increasingly used as a tool for managing research projects. We see new (old) actors such as librarians or administrative staff entering the scene. Metadata are controlled and even enhanced, which implies a gain in consistency, and the validity of the document is checked or controlled through the workflow.

Along with these institutional repositories we observe an increasing number of portals or websites offering a federated research, either for all types of scientific output, or dedicated to ETDs.

- On a worldwide level OAIster* provides access to electronic theses deposited in archives compliant with the OAI protocol for metadata harvesting the NDLTD Union Catalogue does likewise.
- The NARCIS* portal in the Netherlands receives input from all Dutch universities as well as from national research organizations. Its subset "Promise of science" is dedicated to doctoral theses.
- The Scandinavian portal DIVA* references doctoral and even more Master theses from 24 participating colleges and universities.
- EThOS* (Electronic Theses Online System) as a project funded by JISC (Joint Information Systems Committee) and RLUK (Research Libraries UK), opened in 2009 in its Beta version. It's a "one stop shop" for secure access to research theses, simultaneously increasing the visibility of UK Higher education postgraduate research, and providing additional services such as print on demand or digitization of paper documents.
- In France, PASTEL*: theses from ParisTech (Paris Institute of Technology) includes 12 engineering schools. Besides providing free access to

ETDs, ParisTech developed since 2003 the project "ParisTech Graduate School"[8] with access to open courseware.

- DART Europe* is a partnership of research libraries in 14 European countries supported by LIBER* (Association of European Research Libraries). DART-Europe is the European Working Group of NDLTD* and allows for search of 110,000 full text theses.
- Since 2004, about 40 Canadian universities deposit metadata and electronic theses in the Theses Canada Portal*[9]. Moreover, theses and dissertations digitized between 1998 and 2002 are freely accessed. In 2008: it contained 300,000 ETDs on microform and 50,000 are also available electronically.
- The Australasian Digital Theses Program*[10] was funded by the Australian Research Council (ARC). The document format standard is PDF and the portal is OAI-PMH compliant. In 2009, 25,000 digital theses out of 150,700 from 41 universities are available in electronic format.

7.4.2 Quality issues and preservation

A viable website should at the same time ensure quality issues and offer long-term digital preservation.

Archiving and preservation aspects are treated with various approaches in terms of technologies used.

In Europe, DRIVER* "a pan-European infrastructure for digital repositories" (Robinson, M. et al. 2009) emphasizes the need for targeted, refined and standardized harvesting. Thus, DRIVER recommends the following qualifications for "type" element:

"info:eu-repo/semantics/bachelorThesis",
"info:eu-repo/semantics/masterThesis",
"info:eu-repo/semantics/doctoralThesis (Bologna Convention) "

instead of various qualifications used :

"Cranfield: *<dc:type>Thesis or dissertation</dc:type>*
<dc:type>Doctoral</dc:type>
<dc:type>PhD</dc:type>
DIVA: *<dc:type>text.thesis.doctoral</dc:type>*
Humboldt: *<dc:type>Text</dc:type>*
<dc:type>dissertation</dc:type>"

(Robinson, M. et al. 2009)

France developed a standard called TEF* (AFNOR* recommendation since 2006) to doctoral theses. TEF* include a FRBR (Functional Requirements for Bibliographic Records) data representation and "defines a set of preservation metadata

8 ParisTech Graduate School, http://graduateschool.paristech.org/
9 Theses Canada Portal, http://www.collectionscanada.gc.ca/thesescanada/index-e.html
10 Australasian Digital Theses Program, http://adt.caul.edu.au

that will permit long-term thesis preservation" (Boudia, D. et al. 2005) The preservation is maintained by PAC* (Archive Platform at CINES*) based on ISO standard 14721.

At the international level, NDLTD* with the "MetaArchive cooperative project*: digital preservation" uses LOCKSS[11] (ETD 2009)

With regard to the quality of information, the risk of plagiarism has become an important obstacle to the deposit of ETDs in open archives (Davis et al. 2007), a phenomenon which increased with copy and paste facilities and the use of machine translation software.

However, plagiarism is not a recent problem and can be « solved » before deposit. A report by the Commission Ethique-Plagiat funded by the University of Geneva "La relation éthique-plagiat dans la réalisation des travaux personnels par les étudiants"(2008)[12] underlines the importance of training students on document retrieval, and the role of evaluation and control by the authorities (teaching staff). The study recommends the use of software to detect plagiarism.

The attribution of a date stamp during the deposit proves the priority of the work.

7.4.3 Access to full text and confidentiality

Many academic and research institutions have defined policies, declaring the deposit of scientific works in their institutional (or national) repository a mandatory step for either the evaluation of a researcher or the defence of a doctoral dissertation. Academic regulations exist since 2005 for masters and PhD students at the University of Edinburgh[13] and since 2006 at the University of Liège (Belgium)[14] and at Leiden University (Netherlands)[15]

The need for restricted access to parts of the documents may create obstacles to these policies, but different kinds of solutions have been developed. Indeed a mandatory deposit does not necessarily include the authorization for worldwide dissemination of the full text. Aspects like confidentiality or quality criteria may lead to a restricted access to the document itself, e.g. through an intranet or on the campus, while the metadata are available to everyone. Users of such repositories find the information at different levels or stages of their visit.

Partial access to the full text: Authors may be charged with the submission process only once. What procedure to adopt if the dissertation is declared confi-

11 LOCKSS (Lots of Copies Keep Stuff Safe), http://www.lockss.org/lockss/Home

12 La relation éthique-plagiat dans la réalisation des travaux personnels par les étudiants, Bergadaà, M., Dell'Ambrogio, P., Falquet, G., McAdam, D., Paraya, D., Scariati, R., Genève, 8 avril 2008 : http://responsable.unige.ch/rapportunige/RapportPlagiat_Unige2008.pdf

13 University of Edinburgh, http://www.era.lib.ed.ac.uk

14 University of Liège, http://bictel.ulg.ac.be/presentation_2.html

15 University of Leiden, http://www.research.leiden.edu/phd/

dential in part or in its entirety, if it is based on articles for which restrictions exist or when the moving wall applies?

On the technical level software platforms allow the author to deposit the document once, but separated into different files, and to declare some parts under embargo. A user from outside the campus may come across a doctoral thesis for which the table of contents, introduction and conclusion as well as part 1 and 2 are freely accessible, but the link given for the other parts won't work for him. The University of Leiden has adopted this practise for its repository, including even the date of the end of embargo.

Finland offers a different solution. Many Finnish Doctoral theses are based on articles. The full text documents accessible through the repository list the articles in the table of contents and include a summary in the corresponding chapter. The publications seem to be joined as appendices to the thesis, but are not included in the open access version.

A third way to deal with restricted parts of ETDs is shown by the University of Oslo repository DUO*. The bibliographic record includes the list of papers with a hyperlink to the commercial publisher. So the end-user may access these parts if he has a subscription to the journal or if he's willing to pay for it (example: http://wo.uio.no/as/WebObjects/theses.woa/wo/0.3.9)

Records without theses: An increasing number of repositories contain bibliographic records without full text. To support authors with their deposits and to increase their willingness to participate, many repository administrators have started to upload citations from external sources. In some cases the metadata for ETDs originate from library catalogues, to be completed by the author. The absence of the full text is then a phenomenon limited in time. Duplicate entries in repositories may be generated when metadata are added in the deposit process and are uploaded from a bibliographic file. For other universities the institutional repository should reflect the scientific production. Comprehensiveness of bibliographic records prevails over access to the documents.

How are the users informed about the presence/absence of full text documents? Again different solutions can be observed (for details see Stock 2008):

- A check box requesting "full text results only" in the search menu
- Use of icons in the list of results
- Information in the record display
- On specific pages dedicated to a community, a department, etc.

Other solutions: The Imperial College London Repository has chosen to give public access only to validated full-text dissertations, while an in-house platform is used for the workflow and administrative purposes (Jones 2007)

In France all doctoral theses must be referenced in the national bibliography, whereas access to the full text (paper or electronic form) may be "confidential". The present metadata scheme TEF* (TEF 2007) allows to identify and exclude passages not to be disseminated, and anticipates the co-existence of a complete version and a public version of the document.

What happens if the repositories are harvested by service providers?

Several checks with OAIster* showed that in some cases the service succeeds in eliminating OAI records without full text (e.g. Lund), while not in other cases. Hasselt University exposes to metadata harvesting only those documents for which the full text is available (Goovaerts 2007).

7.4.4 Masters theses and other student's works

PhD theses are often subject to academic regulations concerning their dissemination. Ideally, the degree is only obtained when the document is published in print or electronic form. Master's theses are far less controlled by legal dispositions. Conservation of the paper copy in the local library is not always guaranteed and consultation by other students may be subject to authorization.

Disseminating master's theses through repositories is a common occurrence in northern countries. The Scandinavian portal DiVA* provides access to theses at different levels, from master thesis to first term paper. The question whether or not to include master's theses in open archives gave way to ardent exchanges on German discussion lists in the past and is not clear cut in France either. Repositories are seen as a showcase for scientific output, and student's works on a master or even bachelor level don't count as scientific publications, therefore don't belong in this category.

Dissemination of master theses may follow different objectives: making one's work known to fellow students and subsequent as well as to alert future employers. Thus THESA*'s references of cutting edge theses subjects are geared to the economic world, whereas "DUMAS - Dépôt Universitaire de Mémoires Après Soutenance"[16] aims to increase the visibility of master theses as well as the teaching activity of the universities. Others like (mémSIC[17]) instead make a selection of the best works.

7.5 Outlook

The landscape of online access to theses has changed in many ways over the past years. Digitizing paper copies was re-placed by workflows covering every stage of the production and dissemination of a thesis, including metadata and quality issues. Access through a list of titles on a webpage gave way to portals regrouping multiple repositories. One of the major actors in this process is NDLTD.

Technical developments have greatly facilitated the dissemination of ETDs, increasing their visibility to a worldwide level. However, legal issues especially copyrighted parts of third authors used in the thesis became an obstacle. Different solutions have been found and adopted so far. Indeed the growing number of the-

16 DUMAS - Dépôt Universitaire de Mémoires Après Soutenance, http://dumas.ccsd.cnrs.fr/
17 mémSIC, http://memsic.ccsd.cnrs.fr

ses "available" on the internet comes with an increasing diversification as to the kind of access available to the full text and to its contents. Repositories mix full text entries with records without documents or theses with only partial access. Institutions alert users about these differences, but not in a consistent way. And, this diversification extends to other students works.

In an earlier study (Stock 2008) we observed important differences between countries with regards to the number of theses available online, as well as to the percentage written in English. Scandinavian countries as well as Belgium (and the Netherlands) are highly tolerant with regards to the language choice for a thesis: 50 to 90 percent of ETDs appear in English. France on the other hand has a low percentage of theses that are not written in French. "Pioneer deposits" of theses in repositories are mostly written in English and seem to indicate the willingness to give the widest access possible to one's work both through the choice of language and through the internet.

The growing complexity of the ETD landscape calls for explicit policies in the future to inform the user of a given repository on its deposit, validation, access and reuse of a thesis. Useful tools like OAIster* or the sites of service providers who harvest their records should be examined with a critical eye, especially when the primary need of the end-user is access to full text.

Glossary

ABES: Agence Bibliographique de l'Enseignement Supérieur (operating agent of the French academic union catalogue and ILL system): http://www.abes.fr/abes/en/index.html

AFNOR: Association Française de Normalisation (French standardisation organisation) http://www.afnor.org/

Australasian Digital Theses Program: http://adt.caul.edu.au/

CCSD: Centre pour la Communication Scientifique Directe (CNRS unit) http://www.ccsd.cnrs.fr/?lang=en

CINES: Centre informatique national de l'enseignement supérieur http://www.cines.fr/

CNRS: Centre National de la Recherche Scientifique (French National Research Organisation): http://www.cnrs.fr/index.php

Cyberthèses: thesis electronic archive and diffusion program http://www.cybertheses.org/?q=en/node/32

Cyberdocs: collaborative development site of Cyberdocs platform http://www. cyberdocs.org/en

DART-Europe: E-theses Portal http://www.dart-europe.eu/basic-search.php

DissOnline: German project http://www.dissonline.de/eng/links/index.htm

DRIVER: Digital Repository Infrastructure Vision for European Research http://www.driver-repository.eu/

DIVA: Digitala Vetenskapliga Arkivet - Academic Archive On-line http://www.diva-portal.org

DUO: Digitale utgivelser ved UiO http://www.duo.uio.no/englishindex.html

EThOS: Electronic Theses Online System http://www.ethos.ac.uk/

HAL: HAL - Hyper Article en Ligne (HAL - Hyper Article on Line): http://hal.archives-ouvertes.fr/index.php?langue=en&halsid=13aa66ff4f10ccfb0298c0d5f4ef2860

INIST: Institut de l'Information Scientifique et Technique (CNRS institute for scientific and technical information): http://international.inist.fr/

INRIA: French national institute for research in computer science and control: http://www.inria.fr/index.en.html

INSA Lyon: Institut National des Sciences Appliquées de Lyon http://www.insa-lyon.fr

IRIS: digital library of University of Lille 1 https://iris.univ-lille1.fr/dspace/

LIBER: Association of European Research Libraries http://www.libereurope.eu/

MathDoc: French network for documentation in mathematics and server for the management of ETDs run by the university of Grenoble-1 and the CNRS: http://math-doc.ujf-grenoble.fr/Theses/index-en.php

MetaArchive cooperative: digital preservation http://www.metaarchive.org/

NARCIS: portal of Dutch universities http://www.narcis.info

NDLTD: Networked Digital Library of Theses and Dissertations http://www.ndltd.org/

OAIster: union catalogue of digital resources http://www.oaister.org/

OpenDOAR-the Directory of Open access repositories: http://www.opendoar.org/

OpenSIGLE: System for Information on Grey Literature in Europe: http://opensigle.inist.fr/?locale=en

ORI-OAI: Outil de Référencement et d'Indexation en réseaux de portails compatibles OAI-PMH http://www.ori-oai.org/display/ORIOAI/ORI-OAI.ORG

PAC – Archive Platform at Cines: http://www.cines.fr/spip.php?rubrique152&lang=en

PASTEL: digital archive produced by the Paris Institute of Technology http://pastel.paristech.org/perl/set_lang?langid=en&fromurl=/

ROAR: Registry of Open Access Repositories http://roar.eprints.org/

SIGLE: see OpenSIGLE

STAR: Signalement des thèses, archivage et recherche (referencing, archiving and retrieval of ETDs): http://www.abes.fr/abes/page,428,star.html

TEF: Thèses Electroniques Françaises (Metadata for French e-Theses): http://www.abes.fr/abes/documents/tef/index.html

TEL-HAL: multidisciplinary theses server http://tel.archives-ouvertes.fr/index.php?langue=en&halsid=49akpctkmleu5aejcr4n9onpv2

THESA: THESA provides information on doctoral theses currently under way in the French accredited higher education establishments (Grandes Ecoles) http://thesa.inist.fr/eng/Accueil.htm

Theses Canada Portal: http://www.collectionscanada.gc.ca/thesescanada/index-e.html

UMI: ProQuest UMI Dissertation Publishing: http://www.proquest.com/en-US/products/dissertations/

SUDOC: Système Universitaire de Documentation (academic union catalogue of serials and monographs): http://www.sudoc.abes.fr/DB=2.1/LNG=EN/START_WELCOME

All websites visited in June 2010

Bibliography

Boudia, D., Gómez de Regil, R.M. (2005), TEF: Metadata for French Theses and Dissertations. –In: GL7, The Seventh International Conference on Grey Literature: Open Access to Grey Resources, 5-6 December 2005, Nancy, France. http://opensigle.inist.fr/handle/10068/697806

Davis, P.M., Connolly, M.J.L. (2007), Institutional repositories: evaluating the reasons for non use of Cornell University's Installation of DSpace.- D-Lib Magazine, vol. 13, n.3/4, March/April 2007. –ISSN 1082-9873. http://www.dlib.org/dlib/march07 davis/ 03davis.html

ETD 2009, International Symposium on Electronic Theses and Dissertations, NDLTD, Pittsburgh, PA, USA, June 10-13, 2009, http://www.library.pitt.edu/etd2009/

ETD Guide (2009), http://www.ndltd.org/documentation

Fox, Edward A. et al. (1996), National Digital Library of Theses and Dissertations: A Scalable and Sustainable Approach to Unlock University Resources. – In: D-Lib Magazine, vol. 2, Sept 1996. –ISSN 1082-9873. http://www.dlib.org/dlib/september96/ theses/ 09fox.html

Goovaerts, M. (2007), Submission based on document type: DSpace development at Hasselt University. –In: DSpace User Group Meeting 2007, Food and Agriculture Organization of the United Nations, Rome, Italy. http://www.aepic.it/conf/DSUG2007 viewabstracte8c8.html?id=201&cf=11

Jones, R. (2007), On Integration. –In: DSpace User Group Meeting 2007, Food and Agriculture Organization of the United Nations, Rome, Italy. http://www.aepic.it/conf/ DSUG2007/viewabstract8a32.html?id=200&cf=11

Mermet, J.-M., Prudhomme, B., Joly, M., Pinon, J.-M. (1998), CITHER: un modèle de bibliothèque numérique de mémoires de thèses. - In: Document numérique, vol. 2, no. 3-4 (1998), pp.51-72. ISSN 1279-5127

Nicolas, Y. (2006), TEF : métadonnées pour les thèses électroniques françaises. – In: ETD 2006, The 9th International Symposium on Electronic Theses and Dissertations, June 7 - 10, 2006, Quebec City, Quebec, Canada. http://www6.bibl.ulaval.ca:8080/ etd2006/pages/papers/SP6_Yann_NICOLAS.pdf

Robinson, M., Manghi, P., Horstmann, W., Peters, D., Manola, N., Summann, F. (2009), DRIVER tutorial. –In: CERN workshop on Innovations in Scholarly Communication (OAI6), 17-19 June 2009, University of Geneva .http://indico.cern.ch/contribution Display.py?contribId=40&confId=48321

STAR (2009), Carrefour des thèses électroniques françaises. http://www.abes.fr/abes/ DesktopDefault.aspx?tabindex=3&tabid=428

Stock, C. (2008), Open access to full text and ETDs in Europe: improving accessibility through the choice of language? - In: GL9, The Ninth International Conference on Grey Literature: Grey Foundations in Information Landscape, 10-11 December 2007, Antwerp, Belgium. http://opensigle.inist.fr/handle/10068/697889

TEF (2007), Métadonnées des thèses françaises. http://www.abes.fr/abes/documents/ tef/index.html

UMI (2009), http://www.umi.com/en-US/products/dissertations/

All websites visited in June 2010

Chapter 8

Grey Documents in Open Archives

Daniela Luzi, National Research Council, Italy

8.1 Introduction

Science is a social activity [Merton, 1973] based on a cumulative process of knowledge building and sharing. This process relies on the efficacy of an information infrastructure defined as "the technological, social, and political framework that encompasses the people, technology, tools, and services used to facilitate the distributed, collaborative use of content over time and distance" [Borgman, 2007]. The use of ICT (Information and Communication Technologies), and of Internet in particular, is radically changing the way in which scientific research is carried out and consequently the way in which information is shared and exchanged. This has called into question the roles and functions of the principal actors that, in the so called 'print era', contributed to adding value to the process of scientific communication and, at the same time, solicit new petitions for the free circulation of knowledge proposed by the Open Access movement.

Many of these changes are still ongoing. They involve both the use of the new technologies and a cultural shift in the information production and dissemination practices among different scientific communities. A slow and still to be fully negotiated reappraisal of the roles of the different actors in the value chain of scholarly communication is emerging, starting from that of the commercial publishers who propose hybrid models (author pay, institution pay, pay per view) in search of alternatives that do not weaken their income position. Lastly, at the institutional level, both nationally and internationally, a gradual establishment of the principles of free knowledge circulation (starting from those stated in the Berlin Declaration) is underway as it is the proposal of policies, the implementation of which could set the direction of and thus substantially accelerate these changes (take, for example the policies adopted by several national and international organizations to make submission of scientific publications mandatory).

Nevertheless, the conviction is now emerging that ICT is still not being exploited to the full as it has so far merely telematically duplicated the conventional process of scholarly communication; just as the paper-based documents have simply been transformed into analogous digital versions. Projects and research in this sector (Web 2, OAI-ORE) and examples of applications in advanced sectors

are now proposing new and promising solutions pointing to possible much more radical changes in the way research is carried out, how information is accessed and then re-used. The challenge lies not so much in rendering all research products accessible but reconstructing the links among the various scientific outputs, thus reproducing the individual phases of the process of scientific enquiry [Van de Sompel et al. 2009].

If these are the premises, outlined here concisely and in a deliberately simplified way, what is the position of Grey Literature (GL) today? There is no doubt, as many have claimed [Gelfand 2005; Banks et al. 2007], that GL is at *home* in open archives, whether in the form of e-print archives or Institutional Repositories (IRs). Indeed, precisely insofar as scientific GL records and documents the results of the various research phases, its inclusion in the open archives legitimizes it as a scientific artifact and acknowledges its added value. In a scenario of interconnections among the various research products and of the integration between services and information sources, this means that GL has every right to be included in the process of the production and transmission of knowledge worthy of being conserved and disseminated.

8.2 GL in scholarly communication

In the print era, GL acted as a channel for the dissemination of scientific information that ran parallel with that of commercial publishing and was not subject to set production and dissemination rules. This lent GL the characteristics of an informal communication which, according to Meadows [Meadows 1998] is "often ephemeral and made available to a restricted audience only". The similarity between this definition and several of the characteristics peculiar to GL, often cited in support of Wood's definition [Wood 1982, Auger, 1993], is quite apparent.

The opposite pole is represented by formal communication, which culminates in a publication "available over long periods of time to an extended audience" [Meadows 1998]. However, there has never been a clear-cut distinction between the two (and the development of GL is evidence of this) and it is currently dwindling even further in the context of digital media and networked communication. For this reason, in the wider acceptance of the term 'scholarly communication' [Borgman, 2007] defined as the study of "formal and informal activities associated with the use and dissemination of information through public and private channels", GL can provide a good vantage point from which to observe the continuum – whether pre-publication [Harnad 1990] or electronic publication [Kling, 1999, Borgman 2000] – of information exchange activities underlying scientific inquiry.

As part of the renewed interest shown in the ongoing changes in scholarly communication, several recent studies [Pepe et al, 2009] made use of the technique of descriptive laboratory accounts of science–in-action in order to analyse the various artifacts produced by specific scientific communities in the different phases of the life cycle of research activities. These studies have confirmed that in

each phase numerous artifacts are produced, each with its own role, specific content and a different communicative function. They also highlighted their continual, progressive updating and enhancement by which they are transformed into new products, ultimately arriving at the "formal" artifact that can be condensed into a publication.

Van de Sompel [Van de Sompel 2004] describes these artifacts as units of scholarly communication that "reflect the changing nature of the information assets produced and consumed in the scholarly endeavors". Borgman [Borgman, 2007] describes the continuity of scholarly communication, pointing out the gradual shaping and reshaping of information intended for specific recipients. Taking these considerations as a starting point, GL can be described on the basis of the functions it performs within scholarly communication. Thus – despite a degree of overlap and hazy outlines – report literature accomplishes the function of describing detailed results of specific research phases, while conference papers and preprints (generally *reshaped* from report literature) are focused on communicating them to a specific audience. In the context of higher education, theses perform the function of certifying the acquisition of an academic qualification but at the same time report the results of the experimental research undertaken by the candidate or else describe the state of the art of a specific topic, whereas courseware materials, which are also developed as a function of the academic level of the reference audience, provide a systematic framework for the knowledge so far gained in a given disciplinary field.

Side by side with these artifacts, modes of web-based interaction have grown up that increasingly resemble face-to-face communication and the collaborative construction of knowledge bases (the well known and extensively consulted wikis). Starting from the first bulletin boards, nowadays several blogs (for example those following the Open Access debate) embody one of the many tools used by virtual communities of scholars to share open peer commentary and documents. From the point of view of circulation, blogs sometimes become actual sources of information that are used to keep up to date on the developments in certain topics and to follow the progress of the ongoing discussion. These tools open up new horizons for scholarly communication (and also for other communications) and give form to constantly growing and transformed grey information or grey content, which deserves a more detailed and specific analysis.

8.3 From report literature to scientific artifacts

GL develops in practically all disciplines in response to a need for scientific communication, ranging from that of providing a detailed documentation of research results, without the space limitations of journal articles, to that of reducing the time between information diffusion and their actual publication. These needs have formed the basis of many of the initiatives aimed at setting up open access archives that have often viewed GL documents as a test bench for trying out new

dissemination methods and have generally involved the same actors as those who contributed to the development of GL.

The informal nature of GL, previously considered to be one of its main limitations, has actually allowed for both a transformation and a constant increase in the number of document types which corresponded to specific information needs depending on the disciplinary context involved. As mentioned above, each document type usually contains specific information (although variations in the GL field area are indeed the rule). Report literature represents the basic nucleus of paper-based GL, the various types of technical papers (interim reports, research memoranda, working papers) generally describe a particular phase of the research activity and report the results, providing a detailed documentation of data and processing, as well as of the procedures and methods used to analyse them. The absence of publishing rules restricting the length of the articles makes it possible to give a detailed description of the data and procedures adopted. These documents often represent the only information source in which it is possible to find these data that are necessary to duplicate and verify the research objectives and results, and possibly draw new insights from them.

What in the print era made up the content of technical reports or the annexes thereto is now accessible in the form of "compound units" [Van de Sompel et al. 2007], which is also denoted as "datument, a compound document where all the compounds (data, text, software images, links) are part of the whole" [Murray-Rust 2008]. The transposition of these documents into a digital environment allows their information content to be enhanced considerably. It offers the opportunity of including in the electronic document hyperlinks not only to other bibliographic sources, but also to other types of scientific artifacts, such as simulations, videos, data sets, original lab notebooks and even software used to display and/or further process such data. In this way both research results and the process by which they have been obtained are *reproduced*.

In data-intensive and highly collaborative disciplinary fields (such as molecular biology, earth and space sciences, but also in the social sciences) access to these data become essential. This certainly raises questions related to copyright, authorship and access licenses (and the OA movement has in recent years included in its own agenda also the issues of free access and re-use of datasets), as well as the need to tackle problems linked to the storage and retrieval of this information.

If the digital document is transformed into a compound document, it is necessary not only to find suitable techniques of dataset retrieval from which one or more documents may be prepared, but also link together the various different document types, ranging from technical reports, degree theses, and papers delivered at congresses, down to articles published in commercial journals, each of which accessible from different network locations. The link between the various documents and the data on which they are based means that each scientific artifact serves as an entry point to the set of related artifacts [Pepe et al. 2009]. This is useful for reconstructing the entire research process as well ad capturing the progressive shaping and reshaping of scholarly communication.

Initiatives in this sector are found both at the theoretical level and in systems

developed in specific disciplinary fields. The former are being developed by the Open Archives Initiative – Object Reuse and Exchange protocol [OAI-ORE 2008] which published the specifications for handling aggregations of compound information objects of web resources in 2008.

Several systems, which would warrant a separate treatment, point to a growing number of web sites run by research organizations but also by commercial publishers. Here only a few examples will be given. In the field of biomedical and life sciences a number of examples exist, including Nature Proceedings [Nature Precedings], a free service produced by the publishers of *Nature,* which collects pre-publication research and preliminary findings in the biomedical and life sciences. Also within life sciences, BioLit [BioLit Project; Fink et al 2007], set up by the University of California, supplements the articles published by the Public Library of Science (PLoS) with the information contained in the Protein Data Bank (PDB), while another archive, SciVee allows the open uploading of published articles accompanied by the relative video or podcast presentations. In astrophysics mention must be made of the NASA funded Smithsonian Astrophysics Data Systems (ADS) which collects both astronomic and physics literature and link it to data collected by space missions and ground-based observations [SAO/NASA ADS; Eichhorn et al. 2006]. Lastly, also in the social sciences, the Council of European Social Science makes the CESSDA archive available, which allows the retrieval of the datasets and variables of sociological surveys, longitudinal studies, census data collected in the various European countries.

8.4 From the preprint culture to Open Access

Preprint culture has generally been attributed to specific scientific communities, in the first instance of physicists and computer scientists. In actual fact, many studies, beginning with those of Gavey and Griffith [Garvey et al. 1967] in the field of psychology, indicate that above all in those sectors in which no short-term commercial applications are to be expected, there is a widespread attitude among scholars to consult and exchange preprints or other types of document that have not yet been formally published in journal articles, in order to keep abreast of the latest research developments and to seek comments via private circulation to friendly reviewers.

The physics sector may be considered emblematic for various reasons. It must not indeed be overlooked that physicists were among the first users of networks even before the advent of the Internet. The World Wide Web was conceived at the European Organization for Nuclear Research (CERN), and the first US web site was opened a few months later (December 1991) at the Stanford Linear Accelerator Center (SLAC) precisely to gain remote access to the SPIRES-HEP (Stanford Public Information Retrieval System-High Energy Physics) database, one of the richest electronic archives set up to handle preprints.

The SPIRES database is a good example of the tight collaboration network that exists among libraries, as well as the productive interaction between libraries and their own local and remote users. This database was developed to facilitate the distribution of preprint lists diffused by post by the SLAC library in the early 1970s and was progressively enhanced thanks to the contributions of both libraries of similar institutions and SLAC researchers, who developed advanced tools to support the timely diffusion of GL documents [Carroll et al. 1994; Kreitz et al. 1996, O'Connel 2000]. These collaborations and synergisms were able to nourish the preprint culture in that they triggered a virtuous circle between the demand for timely information and a wide range of information sources on offer, which simply reinforces the attitudes of sharing and exchanging information strongly.

This is the environment in which the well-known ArXiv archive developed. In 1991 Ginsparg created a centralized system for the electronic distribution of e-prints at Los Alamos National Laboratory. Designed for a group of about 160 High Energy Physics researchers, the future ArXiv e-print archive expanded rapidly. In just a few months it was not only extended to 1000 users (preprint readers and submitters) and adopted in other areas of physics, but was also introduced and used in other disciplines (mathematics, computer science, linguistics and cognitive sciences, and even in economics).

Initially the system was conceived as a central preprint archive, that is, the type of GL document closest to the final versions of a journal article, not yet subject to copyright and not yet subjected to the formal peer-review process. E-prints apparently mark the transition from hard-copy preprint to the electronic preprint. In actual fact the ArXiv archives, and all those based on this model, soon turned into archives in which GL and conventional literature overlapped. For example, in the current version of ArXiv, full-text access of the document is necessary to verify whether, together with the univocal number attributed to the e-print, there is also an indication of the periodical in which it was published. The e-print definition given by the Joint Information Systems Committee (JISC) [Swan et al. 2005] on the one hand highlights its function ("a digital duplicate of an academic research paper that is made available on line as a way of improving access to the paper") and on the other, emphasizes peer review, which becomes the only, albeit important, difference between GL and conventional literature. Currently an unreferred preprint is distinct from a peer-reviewed postprint. Many publishers nowadays also allow authors to self-archive postprints on their own web page or in open archives, considering them *different* from the publisher-generated format.

The main novelty lies in the fact that the ArXiv archive created a self-sufficient communication model, without intermediaries, which enhances and exploits the interactive aspect. So much so that it seems that the phases of acquisition, storage and dissemination coincide and the roles of author and reader, and even of metadata supplier, now overlap.

The self-consistency of this communication model is highlighted also by Van de Sompel [Van de Sompel et al. 2004], who analyses it using the value chain functions as applied by Roosendaal and Guerts [1997] to scholarly communication. Self-archiving corresponds to the *registration* phase which allows claims of

precedence for scholarly funding, the *certification* function, conventionally performed by peer review, is carried out according to Van de Sompel through the procedures of endorsement of potential submitters by peers, but probably also by the reputation of the institutions and by that of its scholars in a scientific community accustomed to collaborating in experiments involving a large number of researchers. The *awareness* function is performed by the dissemination of scientific content that is freely accessible online, by alerting services and by allowing search engines to index content, while the *archiving* function is "based on ensuring adequate redundancy through the operation of a network of separately controlled systems". However, *Rethinking scholarly communication* does not mean giving priority to the ArXiv models over the conventional process of scholarly publishing, but once again designing a flexible and interconnected research infrastructure that can be enhanced by the contributions received from different services and actors, capable of reflecting the information needs of the various scientific communities in order to promote the advancement of scientific knowledge.

The prominence to which the Ginsparg system rapidly rose was not dependent solely on the number of e-prints submitted or the number of archive accesses, but was the result of the convergence of many factors. The ArXiv actually developed at a time in which library budget cuts and the soaring subscription costs (the so-called 'journal crisis' and 'permission crisis') revealed the critical weaknesses of the conventional model of scholarly publishing. This came about at the same time as the expansion of Internet, which instead opened up the possibility of timely information dissemination to a potentially unlimited number of users at relatively low cost.

The best practice represented by the ArXiv archive was therefore a point of reference in the development of the Open Access movement. Converging on it were both the needs of free access and timely dissemination of information pursued by the scientific community and those of libraries, whose role of selection, acquisition and conservation appeared to be strongly limited by the acquisition of bundled packages of journals. Since the Santa Fe convention [Van de Sompel et al 2000], the OA movement took up a proactive stance, identifying organizational structures (service and data providers) and technical instruments (from open source software to the development of the specifications of the Protocol for Metadata harvesting (PMH), enabling archive interoperability) in order to propose an "open scholarly publication framework on which both free and commercial layers can be established". By successively promoting a Green Road (self-publishing by depositing articles in open archives) and a Golden Road (creation of open access journals), the OAI outlined a process of development of scholarly communication in which scholars and libraries repossessed a part of the research products, thus becoming the direct managers thereof. In this way they contribute to counterbalancing a market dominated by the hegemony of a small number of publishers.

The ArXiv currently contains more than 550,000 e-prints, with an annual increase of about 55,000 documents and continues to be a system that is "scientist driven: articles are deposited by researchers when they choose – either prior to, simultaneous with, or post peer review" [Ginsparg 2007]. About 90% of High

Energy Physics preprints are immediately and freely accessible online [SCOAP 2007].

The large number of archives based on the Ginsparg model (including the historical RePec in economics, Cogprint for the neurosciences and E-LIS in information science), although not attaining the same percentages as the HEP preprints, continues to display a constant increase in the number of documents. So much so that many [Borgman 2007; Swan 2005] have pointed out that authors showed a greater propensity to submit their work to these thematic archives rather than self-archiving their works in Institutional Repositories (IRs). The impact of these archives is measured in terms of access statistics, document downloads, and by the list of most cited preprints. A growing number of studies [Harnad et al 2004, Soong 2009] also indicates that open access papers are read and cited more frequently as they are freely and more rapidly accessible. Lastly, the preprint archives are provided with a whole series of gateways that facilitate research in a number of preprint archives (for example, E-Print Network set up by the US Department of Energy) or autonomous systems of citation indexing, such as CiteSeer. These initiatives help indicate alternatives to the conventional process of publication, allowing access at many more points and providing parallel services that enhance the scholarly value chain.

8.5 From institutional repositories to the interconnected knowledge network

While the creation of e-print archives may be considered as a bottom-up initiative, managed directly by a specific scientific community, that of the Institutional Repositories (IRs) marks the official commitment of the universities and research institutions to making their own scientific artifacts freely available. The current tendency is to classify open archives as disciplinary or thematic (the original e-print archives) and the repositories as institutional /departmental, governmental or aggregating IRs [OpenDOAR]. The distinction is made herein for chronological purposes, as IRs derive from the former and above all because from the standpoint of GL, IRs provide a *natural home* for GL. This is due to a series of factors.

The commitment of the institutions has both a political and an operational value. The first aspect is apparent in the commitment to the OA movement as formally expressed the research institutions' endorsement of the Berlin Declaration. This commitment expresses the institutions' intention to regain a proactive role in scholarly communication. Lynch actually views IRs "as a new strategy that allows universities to apply serious, systematic leverage to accelerate changes taking place in scholarship and scholarly communication" [Lynch 2003]. A similar, and even more radical, stance is taken by Crow [Crow 200] in the oft-cited SPARC (Scholarly Publishing and Academic Resources) position paper stressing the role of IRs as an instrument that "increases competition and reduces the monopoly power of journals, and brings economic relief and heightened relevance to

the institutions and libraries that support them". In this way, IRs become also "a digital version of the traditional university press" [Swan et al 2005], the beneficiaries of which, in addition to the scientific community and of course the public at large, are also the libraries.

The aim of the IRs is to develop "a set of services that the university offers to the members of its community for the management and dissemination of digital materials created by the institutions and its community members" [Lynch 2003]. IRs thus represent a comprehensive showcase of the scientific, teaching and cultural activities of a scientific institution. Therefore, the digital materials that are to be rendered freely available include all types of research products (preprints, postprints, theses, conference papers, monographs, research data sets and databases), teaching materials (courseware, lecture notes, etc.) and of course new kinds of grey contents. Each scientific community can decide what kind of collections are to be self-archived, the relative format (whether full-text or bibliographic references) and can ultimately indicate the rules governing access (for instance, by restricting full-text accessibility exclusively to members of the institution).

The organization of scientific content into collections ensures greater visibility of GL, places it in the context of the other artifacts produced by a given community and allows each GL document type to be linked to the appropriate metadata that will facilitate access to and cross-searching among the various repositories and/or search engines. The latest survey of the EU sponsored Digital Repository Infrastructure Vision for European research Project (DRIVER) [Vernooy-Gerritsen et. al 2009] reports that 62% of European IRs consist of GL, 39% of which are theses, 14% proceedings and 9% working papers, compared with 34% journal articles and 4% books and book chapters. These figures aroused considerable interest in that the survey claimed that one of the functions of IRs is to become "a source for grey literature" for users, as well as an "alternative route to toll-access literature" [Vernooy-Gerritsen et. al 2009]. Furthermore, also the figures given in the Directory of Open Access Repository [openDOAR], indicate that of the 1532 IRs surveyed, 50% contain theses and dissertations, 41% unpublished reports and working papers, 15% learning objects, and 4% data sets.

The organizational model put in place by IRs envisages a more central role for libraries, compared with the one played in the original e-print archives. Indeed the library is called upon to participate both in the repository design phase, in which librarianship contributes to defining the collections and to identifying the metadata required to describe them. This is the case both at the stage in which the repository becomes operational and library expertise is needed to validate the data and to identify the long-term strategies for the preservation of digital formats and/or for any digitization of previous collections of paper-based documents. The library is also assigned the tasks of assisting researchers in self-archiving operations and of providing support for the activities of open access advocacy. This is a role that entails ever closer collaboration between the libraries and the reference scientific community, at the same time projecting the latter towards a larger network environment.

No direct link has ever been found between IR development and the previous experience of libraries in handling catalogues and archives explicitly dedicated to the collection of GL [Di Cesare 2006] or their participation in international initiatives aimed at GL dissemination (a typical example is the SIGLE database). There is no doubt that IRs benefit from the previous development of the digital libraries or OPAC and from the libraries' entire accumulated experience of scientific documentation management and thus it can be inferred that they also benefit from the previous complex management of GL. It would be interesting to investigate whether a *GL management culture* coexisted with the preprint culture. This sometimes emerges in the feasibility studies regarding IRs [Lambert 2006] or in setting up IRs based on existing GL collections [Anderson et al. 2007]. In the current debate on OA, increasing mention is made of GL and in particular of the new forms of GL or grey contents. As well as increasing the visibility of GL, this fully justifies its inclusion in research infrastructure networks.

8.6 Concluding remarks

During the print era, GL provided an informal channel for the dissemination of scientific information thanks to the development of its own acquisition and distribution network, albeit for a limited number of experts. Research institutions that were particularly sensitive to its information value set up specialized services for its collection and dissemination; researchers used their own transmission channels and/or were helped by specialized librarians to 'dig it out'. The latter undertook to identify the most suitable ways and means to index and catalogue it. Lastly, several international initiatives, notably SIGLE, funded programs aimed at supporting cooperation among the European countries in the collection and dissemination of GL.

Nowadays open archives are one of the new channels of GL dissemination that enormously amplify its user basin and situate it deservedly in the continuum of scholarly communication. The coexistence of GL with conventional literature actually provides an *ideal, complete coverage* of the research results of any given scientific institution or disciplinary community. It also enables it to be included in the wider debate on Open Access and in the numerous initiatives being developed around this movement.

The distinction between GL and conventional literature is becoming increasingly blurred: GL, in its various forms, is "made public" [Borgman 2007] on the web at a previously unimaginable speed, and the main difference vis-à-vis conventional literature is inherent mainly in the fact that GL is not subjected to any formal peer-review process. New forms of grey contents, which represent essential sources of information for the advancement of knowledge, are opening up new fields of study concerning the conduct of scientific research, as well as the need to preserve and disseminate these artifacts. In this framework, the experience ac-

quired by GL specialists can make a substantial contribution to facilitating and orientating the constant evolution of scholarly communication.

References

Anderson N., G. Hodge, A. Japzon (2007), Harnessing NASA Goddard's Grey Literature: the Power of a Repository Framework. - In: Publish Research Quarterly 23 pp. 58-64.

Auger, C. P. (1993), Information Sources in Grey Literature, 3rd ed. London: Bowkwe-Saur.

Banks, M. A., C. de Blaaij (2007), Implications of copyright evolution for the future of scholarly communication and grey literature. In: Eighth International Conference on Grey Literature : Harnessing the Power of Grey, 4-5 December 2006 / compiled by D. J. Farace / compiled by D. J. Farace and J Frantzen. – Amsterdam: TextRelease, Feb. 2007, pp 43-52.

Berlin Declaration (2003), The original Berlin Declaration is located at <http:// www.zim.mpg.de/openaccess-berlin/berlindeclaration.html>.

Borgman C. L. (2000), Digital Libraries and the Continuum of Scholarly Communication. – In: Journal of Documentation 56 (4) July 2000, pp- 412-430.

Borgman, C. L. (2007), Scholarship in the digital age: information, infrastructure, and the Internet. Cambridge: the MIT Press.

Carroll B. C., G. A. Cotter (1994), A new generation of grey literature: the impact of advanced information technologies. - In: First International Conference on Grey Literature: Weinberg Report 2000, 13-15 December 1993 / compiled by D. J. Farace. – Amsterdam: Amsterdam: TransAtlantic, 1994, pp 5-17.

Crow, R. (2002). The Case for Institutional Repositories: A SPARC Position Paper, Washington: SPARC, 37 p. http://www.arl.org/sparc/publications/papers/index.html.

Di Cesare, R. D. Luzi, R. Ruggieri (2006), Open Archives and SIGLE participation in Italy: Is there a subtle link between the two experiences? - In: An International Journal on Grey Literature 2 (1) spring 2006.

Eichhorn, G., A. Accomazzi, C. S. Grant, E. A. Henneken, D. M. Thompson, M. J. Kurtz, S. S. Murray (2006), Connectivity in the Astronomy Digital Library, ArXiv preprint. arXiv:cs/0610008. http://arxiv.org/abs/cs/0610008. To appear in Library and Information Systems in Astronomy V.

Fink, L. L., P. E. Bourne, (2007), Reinventing Scholarly Communication for the Electronic Age. - In: CTWatch Quarterly 3 (3) Aug 2007, pp. 26-31. http://www.ctwatch.org/ quarterly/.

Garvey, W. D., B. C. Griffith (1967), Scientific Communication as a social system. – In: Science 146 (3652) 1967, pp. 1011-1016.

Gelfand, J. (2005), "Knock, Knock:" Are Institutional Repositories a Home for Grey Literature? - In: Sixth International Conference on Grey Literature: Work on Grey in Progress, 6-7 December 2004 / D. J. Farace and J Frantzen. – Amsterdam: TextRelease, Jan. 2005, pp10-15.

Ginsparg, P. (2007), Next-Generation Implications of Open Access. CTWatch Quarterly 3 (3) Aug 2007, pp. 11-18. http://www.ctwatch.org/quarterly/.

Harnad S. (1990), Scholarly Skywriting and the Prepublication Continuum of Scientific Inquiry. - In: Psychological Science 1, pp 342- 342, http://cogprints.org/1581/0/harnad90.skywriting.html.

Harnad, S. T. Brody, (2004), Comparing the Impact of Open Access (OA) vs. Non-OA Articles in the Same Journals. - In: D-Lib Magazine 10 (6) June 2004. http://www.dlib.org/dlib/june04/harnad/06harnad.html.

Kreitz, P. A., L. Addis, H. Galic, T. Johnson (1997), Collaborative International Control of High-Energy Physics Pre-Prints. - In: Second International Conference on Grey Literature: Grey Exploitations in the 21st Century, 2-3 November 1995 compiled by D. J. Farace. – Amsterdam: Amsterdam: TransAtlantic, 1996, pp. 33-41.

Kling, R., G. McKim (1999), Scholarly Communication and the Continuum of Electronic Publishing. - In: Journal of the American Society for Information Science, 50 (10) 1999, pp. 890 – 906.

Lambert S., B. Matthews, C: Jones (2006, Grey literature, institutional repositories and the organizational context. – In: Seventh International Conference on Grey Literature : Open Access to Grey Resources, 5-6 December 2005 compiled by D. J. Farace and J Frantzen. – Amsterdam: TextRelease.

Luzi, D. (1996), Internet as a new distribution channel of scientific Grey Literature. The case of Italian WWW servers. - In: Second International Conference on Grey Literature: Grey Exploitations in the 21st Century, 2-3 November 1995 compiled by D. J. Farace. – Amsterdam: Amsterdam: TransAtlantic, 1997, pp 42-55.

Lynch C. A. (2003), Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age. ARL Bimonthly Report 226 Feb 2003.

Meadows, A. J. (1998), Communication Research, San Diego, CA: Academic Press.

Merton, R. K. (1973), The Sociology of Science: Theoretical and Empirical Investigations. Chicago: University of Chicago Press.

Murray-Rust, P. (2008), Open Data in Science. – Nature Precedings http://precedings.nature.com/documents/1526/version/1. Published in Serials Review, March 2008.

O'Connel H. B. (2000), Physicists thriving with Paperless Publishing. ArXiv preprint. ArXiv:physics/0007040v2. http://arxiv.org/abs/physics/0007040. Invited talk at the AAAS Meeting, February 2000 in Washington, DC.

Open Archives Initatives – Object Reuse and Exchange protocol (OAI-ORE), (2008), Ore Specification and User Guide, http://www.openarchives.org/ore/1.0/toc.

Pepe A., M. Mayernik, C. L. Borgman, H. Van de Sompel (2009), From Artefacts to Aggregations: Modelling Scientific Life Cycles on the Semantic Web. ArXiv preprint. arXiv:0906.2549v3, To appear in the Journal of the American Society for Information Science and Technology (JASIST).

Roosendaal, H. P. Geurts (1997), Forces and functions in scientific communication: an analysis of their interplay. - In: CRISP 97, Cooperative Research Information Systems in Physics, http://www.physik.uni;oldenburg.de/conferences/crisp97/roosendaal.html

SCOAP Working Party (2007). Toward Open Access Publishing in High Energy Physics. Executive Summary of the Report of the SCOAP3 Working Party, CERN 2007. http//scoap3.org/files/Scoap3WPReport.pdf.

Smithsonian Astrophisics Data Systems (SAO/NASA ADS), http://adswww.harvard.edu/.

Soong, S. S. Measuring Citation Advantages of Open Accessibility. D-Lib Magazine 15 (11/12) Nov. 2009. http://www.dlib.org/dlib/november09/soong/11soong.html.

Suber, P (2009), News from the Open Access. http://www.earlham.edu/~peters/fos/fosblog.html

Swan A., P. Needham, S. Probets, A. Muir, A. O'Brien, C. Oppenhaim, R. Hardy, F. Rowland, (2005a), Delivery, management and access model for E-prints and Open Access Journals within further and higher education. A Joint report by The Electronic Publishing Innovation Centre in partnership with Key Perspective Limited. Loughborough: EPIC, 121 p. http://eprints.ecs.soton.ac.uk/11001/.

Swan A., P. Needham, S. Probets, A. Muir, C. Oppenhaim, A. O'Brien, R. Hardy, F. Rowland, s. Brown (2005b), Developing a model for e.-print and open access journal content in UK further and higher education. – In: Learned Publishing 18 (1) Jan 2005, pp. 25-40.

Vernooy-Gerritsen, M., G. Pronk, M. van der Graaf (2009), Three Perspectives on the Evolving Infrastructure of Institutional Research Repository in Europe. - In: Adiadne Magazine 59 Apr. 2009 http://www.ariadne.ac.uk/issue59/vernooy-gerritsen-et-al/.

Van de Sompel H., C. Lagoze (2000), The Santa Fe Convention of the Open Archives Initiative. D-Lib Magazine 6 (2) Feb. 2000.

Van de Sompel, H., S. Payette, J. Erickson, C, Lagoze, S. Warner (2004), Rethinking Scholarly Communication: Building the System that Scholars Deserve. - In: D-Lib Magazine, 10 (9) Sep 2004. http://dx.doi.org/10.1045/september2004-vandesompel.

Van de Sompel, H., C. Lagoze C. (2007), Interoperability for the Discovery, Use, and Re-Use of Units of Scholarly Communication. - In: CTWatch Quarterly 3 (3) Aug 2007, pp. 32-41. http://www.ctwatch.org/quarterly/.

Van de Sompel, H., S. Warner, C. Lagoze, R. Sanderson, M. L. Nelson, P. Johnston (2009), Adding eScience Assests to Data Web. - In: Proceedings of the 2nd Workshop on Linked Data on the Web (LDOW2009). ArXiv preprint. arXiv:0906.2135v1; http://arxiv.org/abs/0906.2135.

Wood, D. (1982), Grey literature – the role of the British Lending Division. - In: ASLIB proceedings, 32 (11/12) 1982, pp. 459-465.

Cited web sources

BioLit Project, http://biolit.ucsd.edu/doc/.

CiteseerX - http://citeseerx.ist.psu.edu/.

Council of European Social Science (CESSDA) http://www.cessda.org/.

E-print Network - http://www.osti.gov/eprints/

Nature preceding, http://precedings.nature.com/.

Open Archives Initiative, http://www.openarchives.org.

OpenDOAR - http://www.opendoar.org/.

SAO/NASA ADS - http://adswww.harvard.edu/

Chapter 9

OpenSIGLE - Crossroads for Libraries, Research and Educational Institutions in the Field of Grey Literature

Dominic J. Farace and Jerry Frantzen,
GreyNet International, Netherlands
Christiane Stock and Nathalie Henrot,
INIST-CNRS, France
Joachim Schöpfel, University of Lille 3, France

9.1 Introduction

This chapter is based on a paper[1] presented at the Tenth International Conference on Grey Literature (GL10) in which GreyNet's collections of conference preprints were made accessible via the OpenSIGLE Repository. OpenSIGLE offers a unique distribution channel for European grey literature with roots dating back a quarter century. In the first part of the chapter, the experience of INIST as service provider and GreyNet as data provider will be discussed including recent developments.

Later in the chapter, the draft of a project proposal called for in the final session of that conference will be elaborated. The proposal seeks to explore the capacity required for the OpenSIGLE Repository to develop in multilateral and international cooperation in support of European research infrastructures committed to the open access of grey literature collections and resources. Emphasis is placed on the involvement of libraries, research centers, and institutions of higher education, as well as, requirements for a grey literature network service to sustain further development, exploitation, and promotion of the OpenSIGLE Repository.

9.2 From SIGLE to OpenSIGLE: A Progress Report

SIGLE (System for Information on Grey Literature in Europe) was a unique multidisciplinary database dedicated to grey literature. Up to 15 European partners participated in SIGLE, mostly national libraries or libraries aligned to well-known

1 Farace, D.J., J. Frantzen, C. Stock, N. Henrot, and J. Schöpfel (2009), OpenSIGLE, Home to GreyNet's Research Community and its Grey Literature Collections: Initial Results and a Project Proposal. – In: The Grey Journal : An International journal on Grey Literature, vol. 5, no 1, Spring 2009. ISSN 1574-1796

research institutes. Their principal goals were the centralized collection of scientific and technical reports, theses and other grey material and to facilitate access to these documents through an engagement for document delivery or loan. Created in 1980 and produced from 1984 onwards by EAGLE (European Association for Grey Literature Exploitation), the database was last available through STN International and on CD-ROM via Silverplatter/Ovid until it became dormant in 2005. INIST then decided to make the data publicly available on an open access platform. Details of the migration from SIGLE to OpenSIGLE have been presented at the GL8 Conference[2] held in December 2006 (Schöpfel 2007). And in December 2007, the OpenSIGLE website[3] went live.

This chapter further discusses three related issues dealing with OpenSIGLE:

(1) usage statistics covering two years of access to the repository,
(2) a bilateral cooperative agreement with GreyNet, the Grey Literature Network Service, and
(3) a project proposal exploring the capacity required for the OpenSIGLE Repository to develop in multilateral and international cooperation.

9.2.1 OpenSIGLE Traffic Report

Usage information for a database is at all times interesting for the producer of the information. In this case an additional incentive was the fact that OpenSIGLE records, which migrated from the SIGLE database, had not been updated since 2005. Would then the move to an open access environment be at all "useful" for the grey literature community?

The usage analysis is based on data obtained through phpMyVisites, an open source software for website statistics that works with a javascript image call. Only completely uploaded pages are counted and robots are excluded. The following data provide only a part of the information that can be obtained through phpMyVisites. Other statistics based on server logs might however provide even higher figures.

The first figure shows that the number of visits as well as the number of page views has increased steadily since the opening of the website in 2007. A first peak was reached in July 2008 following a press campaign in the middle of the French holidays. The result is both surprising and rewarding since visits usually decrease during summer months.

The usage of OpenSIGLE continues to increase. In terms of page views and number of visits in which the average duration is 90 seconds, the increase is well over four times the amount in March 2010 compared to March 2009. Visits where

2 Schöpfel, J., C. Stock, and N. Henrot (2007), From SIGLE to OpenSIGLE and beyond: An in-depth look at Resource Migration in the European Context. – In: The Grey Journal : An International journal on Grey Literature, vol. 3, no 1, Spring 2007. ISSN 1574-1796

3 OpenSIGLE - System for Information on Grey Literature in Europe, http://opensigle.inist.fr/

only a single page is viewed represent a stable 50% average of the traffic to the site. These users accessed the database after searching via Google or Google Scholar. While in other cases, users may carry out extensive searches and view hundreds of web pages.

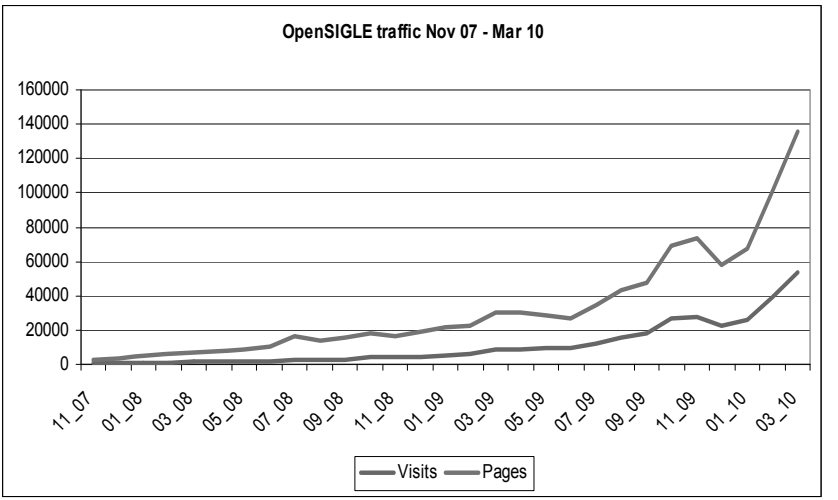

Figure 1: OpenSIGLE traffic report – number of visits and pages viewed

9.2.2 Geographic Origin of Visitors

The software used allows us to monitor the origin of visitors for the top ten countries each month. The sum of 29 months worth of data shows the United Kingdom in the lead, closely followed by the United States. A grouping of other former EAGLE Countries by number of visits to the repository shows Germany, France, Italy, and Spain in their respective order. Countries in the long tail may not appear on a given monthly top ten listing. It is obvious that OpenSIGLE users are not only from Europe, but also from the United States, Canada, and since recently China and Australia. This clearly is an indication that European grey literature presents an interest worldwide.

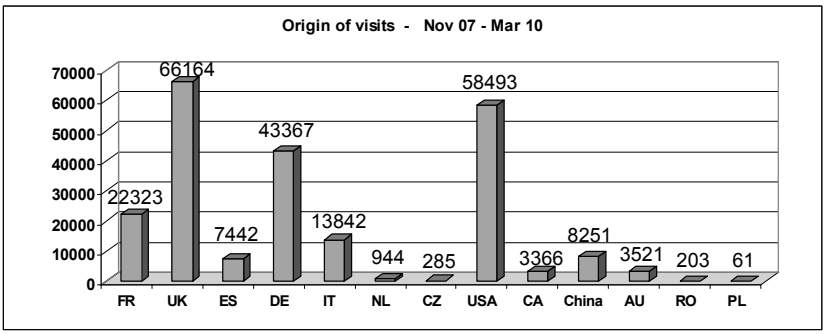

Figure 2: Origin of visitors to OpenSIGLE

9.2.3 Usage and Feedback

Compared to other INIST websites and e-resources, statistics show that 16% to 19% of the users come from North America. OpenSIGLE is in third place among users from this continent preceded by the English version of INIST's institutional website[4] and IndicaSciences[5] - an INIST product dedicated to research evaluation and indicators. INIST websites geared to a French speaking audience receive an average of 7% of the visits from North America.

The analysis of web links as well as feedback through incoming messages reveal that OpenSIGLE is often used in the biomedical and public health sectors. However, at present, statistics do not allow us to go into further detail regarding scientific domains.

During the course of 2008, several requests were received from former users of the STN or Ovid versions of the SIGLE database dealing with complex search strategies. Such questions required another look into the limits of the Jakarta Lucene search engine implemented within DSpace, especially with regard to the length of the search query. It was discovered that Lucene allows for more possibilities than mentioned in the help provided by DSpace. Besides inquiries involving search strategies, users were also interested in the download and export features of OpenSIGLE.

One critical view of OpenSIGLE found on a blog[6], mentions the absence of links to the full text of documents. Of course this is understandable given the fact that it was one of the very reasons why the SIGLE database was discontinued.

4 English version of INIST's institutional website, http://international.inist.fr

5 INIST product dedicated to research evaluation and indicators, http://indicasciences.inist.fr

6 Critical view of OpenSIGLE found on a blog, http://healthinformaticist.wordpress.com/2008/08/28/does-opensigle-exist-for-its-own-sake/

9.2.4 Promotional Activities

Before the official announcement of the launch of OpenSIGLE, the project was presented at a DSpace meeting focused on the exchange of experiences among its users (Grésillaud and Stock, October 2007) [7]. Shortly afterwards, and as a result of that meeting, visitors from Spain and Italy were observed on the OpenSIGLE website. In December 2007, INIST also focused attention on OpenSIGLE during the GL9 Information Walk-Thru at the Ninth International Conference on Grey Literature in Antwerp, Belgium[8].

In May 2008, a short presentation for the French public was given at I-expo (IT conference and exhibit) in Paris. And in July, INIST sent a press release to national and international lists and agencies *i.e*. Information World Review and Research Information. This no doubt resulted in the above mentioned peak of visits in the middle of summer. A brief message about OpenSIGLE was placed simultaneously on the French and international homepages of INIST. Since "news items" are normally less frequent during summer months, the message remained for a longer period of time on these WebPages.

Today OpenSIGLE is indexed by Google and Google Scholar and included in the bookmarks of national libraries and research institutes. Following the creation of the WorldWideScience Alliance and website[9] in June 2008, INIST (a partner in this Alliance) proposed to integrate OpenSIGLE into the WorldWideScience portal. This was realized in September 2008. And, in the web statistics that following month WWS.org appeared as forth partner site for visitors accessing OpenSIGLE through a website with GreyNet.org[10] following closely behind. Overall, these different promotional activities have had a positive impact on the use and branding of OpenSIGLE.

9.3 GreyNet, On the Background and Forefront of OpenSIGLE

Here, the relationship between GreyNet and the former EAGLE Association including its SIGLE database will be addressed. This will then be followed by a conscious positioning of GreyNet in the newfound OpenSIGLE Repository with INIST as its Service Provider.

---

7 Grésillaud, S., and C. Stock (October, 2007), DSpace at INIST-CNRS: one platform, different usages and resulting specific needs/problems. Paper presented at DSpace User Group Meeting 2007, Food and Agriculture Organization of the United Nations, Rome, Italy. Available at http://www.aepic.it/conf/viewabstract.php?id=208&cf=11

8 Grey Foundations in Information Landscape (2007), Ninth International Conference on Grey Literature, 10-11 December 2007 in Antwerp, Belgium. – GL9 Conference Program and Abstracts. – ISBN 978-90-77484-09-8

9 WorldWideScience.org, the global science gateway, http://worldwidescience.org/

10 GreyNet, Grey Literature Network Service, http://www.greynet.org/

In 1992, EAGLE agreed to act as main sponsor for the launch of the International Conference Series on Grey Literature first held in the Amsterdam RAI in December 1993. GreyNet was at that time a newly established network service – driven on two fronts: (1) to promote the field of grey literature and the work of organizations involved in this branch of information the world over, and (2) to stimulate research on grey literature and make the results available both in print and digital (electronic) formats. EAGLE participated as sponsor and/or program committee member in the first five Conferences in the GL-Series.

In early 2005 GreyNet was invited as an observer to the final EAGLE Board meeting at FIZ Karlsruhe upon which the EAGLE Association formally voted to be dissolved. It was at that same meeting that the initial draft of an OpenSIGLE proposal[11] was presented by Dr. Joachim Schöpfel , last in line of EAGLE Presidents.

In the two ensuing years (2005-2007), INIST worked unilaterally on OpenSIGLE, which could then be described as a caretaker repository. In the autumn of 2007, once OpenSIGLE had become operational, GreyNet met with colleagues at INIST to hammer out an agreement that on the one hand would make GreyNet OAI-compliant and on the other hand would expand INIST's role in OpenSIGLE from solely a caretaker to an external service provider. To this end, GreyNet's conference based collections would provide an example of OpenSIGLE's potential for other data providers in the grey literature community.

9.3.1 GreyNet's Collections in OpenSIGLE

In December 2008, five years of research issuing from the GL Conference Series had been uploaded in the OpenSIGLE Repository. The bilateral contact between INIST as service provider and GreyNet as data provider was successful in customizing a metadata record for the enriched publication of conference preprints and the subsequent migration of GreyNet's collections to an open access environment. The bilateral agreement likewise holds for future conferences in the GL-Series, continuing with GL10 records onward.

Retrospective input of the initial four conferences in the GL-Series (1993-1999) would of course make GreyNet's collections comprehensive in OpenSIGLE. To this end, in January 2009, GreyNet purchased from Emerald Group Publishing – former MCB University Press – the rights to allow the full-text papers from the earlier four conferences in the GL-Series to be made available in the OpenSIGLE Repository. This step was not only applauded by the open access community[12], but it also suggests other possibilities to retrieve content controlled

11 Schöpfel, J. (2006), MetaGrey Europe, A Proposal in the Aftermath of EAGLE-SIGLE. – In: GL7 Conference Proceedings, pp. 34-39. – ISBN 90-77484-06-X

12 Posting by Peter Suber on January 29 (2009), http://www.earlham.edu/~peters/fos/2009/01/greynet-buys-rights-to-deposit-papers.html

by commercial publishers[13]. GreyNet proceeded with the production of metadata records, while INIST took on the work of scanning and creating image files for the retrospective records. In October 2009, half of the retrospective input had been achieved.

9.3.2 GreyNet's Potential for OpenSIGLE

The initial reaction from the grey literature community to GreyNet's alliance with OpenSIGLE has been positive; however, due to the brief timeframe in which GreyNet's collections are actually available in the OpenSIGLE Repository, it is too early to provide substantial user statistics. While GreyNet has been receiving monthly reports from INIST generated via OpenSIGLE, GreyNet is looking for other ways to compile use and user statistics via its own channels. In this way, there would be separate data issuing from INIST as service provider and GreyNet as data provider that would allow for comparisons and provide grounds for decision making in the future.

In September 2008, an OpenSIGLE webpage was added to the GreyNet website with hyperlinks to its conference collections already in the repository; and in January 2009 that webpage became a main page on GreyNet's website. Not only did the number of visits to the webpage double in the first half of 2009, but it now also allows for the addition of sub-pages used for promotional and instructive purposes.

The Grey Literature Network Service feels that it has even more to offer OpenSIGLE than its conference collections. Going back to 1992, when GreyNet was first launched, one of its primary goals was to promote the field of grey literature and the work of organizations involved in this branch of information. What EAGLE was to SIGLE, GreyNet could be to OpenSIGLE and more. GreyNet operates internationally and maintains a full-time established network service specializing in grey literature with information products and resources both in print and electronic formats. GreyNet has for the past seven years (2003-2009) often together with colleagues from INIST carried out research projects involving citation analysis, surveys, interviews, as well as standard review of the literature. Over the past years (1992-2009), GreyNet has developed channels for promotional outreach as well as a modest publishing arm. More recently, GreyNet has set up a program of training and instruction in the field of grey literature, which could also be linked to OpenSIGLE. These and other such initiatives would no doubt serve and support future developments in the OpenSIGLE Repository.

13 Posting by Heather Morrison on August 6 (2009), http://www.connotea.org/comments/uri/92b11113ecf827be19a369f21e81161b

9.4 OpenSIGLE Project Proposal, A Feasibility Study

What began unilaterally with the vision and determination of INIST and what has recently been expanded in bilateral cooperation with GreyNet has yet even greater potential for the international grey literature community. GreyNet together with INIST are committed to drafting a project proposal. This proposal will explore the capacity required for the OpenSIGLE Repository to further develop in multilateral and international cooperation in support of European research infrastructures committed to open access of their grey literature collections and resources, where special emphasis is geared to libraries, research centers, and institutions of higher education.

9.4.1 Project Lead-Time

Both INIST and GreyNet have put forth a number considerations and recommendations based on their recent experience with the OpenSIGLE Repository. An inventory of issues and recommendations were collated and will be used in the development of work packages in the implementation phase of the project. Some of the issues include: closing time gaps in record entries, linking to full-text documents as well as plus links to datasets and software, integrating OpenSIGLE in other networks and portals, streamlining the SIGLE Classification scheme, etc.

9.4.2 Project Consortium

Based on the main objective of the proposed project and in relation to the issues that would have to be dealt with in order to achieve this objective, project partners and external advisors need to be identified and brought together in a consortium for the duration of the project. To achieve optimal results, the number of stakeholders in the project would be limited. In the diagram below, the content as well as management base of the project is visualized. However, the names of the prospective organizations, who would be carrying out the projects' roles and tasks are masked here until final confirmation.

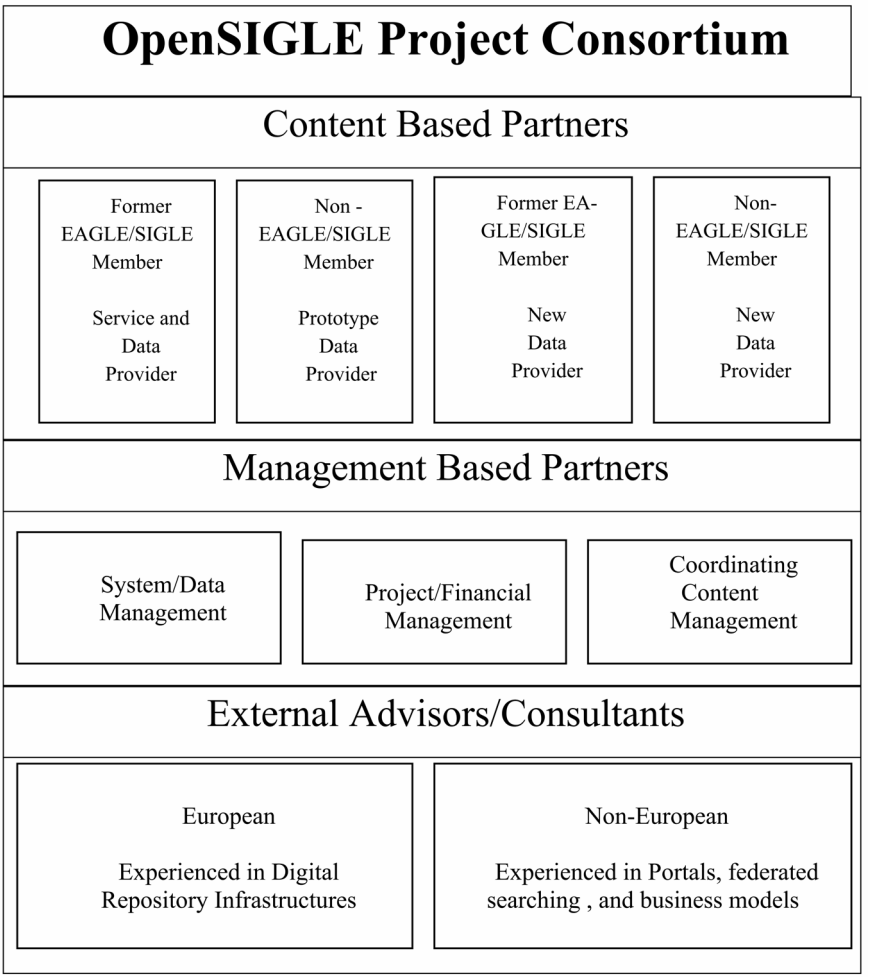

Figure 3: OpenSIGLE Project Consortium

9.4.3 Expected Results and Impact of the Project

This project has its roots in a European framework of cooperation among long-standing infrastructures including national libraries, research centers, and networked services. The outcome of this project would support and strengthen policy development for infrastructures in the field of grey literature, where open access to their collections and other knowledge based resources stand central. The OpenSIGLE Repository with its technical know-how would be sustained by a coordi-

nating infrastructure in the advancement of European cross-disciplinary research well beyond its geographical borders. A draft of this project proposal will be presented during a Panel Session at the Eleventh International Conference on Grey Literature[14] held in the Library of Congress in Washington D.C on 14-15 December 2009 The panel members will take the opportunity to discuss the project proposal in order to illicit feedback from the international grey literature community, raise public awareness to the OpenSIGLE Repository, and solicit leads for further project funding.

14 GL11 Program and Conference Bureau, http://www.textrelease.com/

Appendix: SIGLE – OpenSIGLE Timeline

2009-…	2006-…	2003-…	1999-…	1985-…
2009 GreyNet's Retrospective Collections 1997-1999 in OpenSIGLE; GL11 -Eleventh International Conference on Grey Literature in Washington D.C. with INIST as Co-Sponsor **2008** OpenSIGLE included in WWS.org, World-Wide Science Portal; GreyNet's Collections 2003-2007 accessible via OpenSIGLE; GL10 – Tenth International Conference on Grey Literature in Amsterdam with INIST as Co-Sponsor **2007** OpenSIGLE operational; Bilateral agreement INIST (Service provider) and GreyNet (Data Provider); GL9 – Ninth International Conference on Grey Literature in Belgium with INIST and the EU as Co-Sponsors	**2006** SIGLE migration to OpenSIGLE with INIST as Service Provider; GL8 – Eighth International Conference on Grey Literature in New Orleans with INIST as Co-Sponsor **2005** EAGLE Association dissolved and SIGLE Database dormant; INIST proposal for OpenSIGLE; GL7 – Seventh International Conference on Grey Literature in Nancy, France with INIST as Host and EU as Co-Sponsor **2004** GL6 – Sixth International Conference on Grey Literature in New York, NY with INIST as Co-Sponsor	**2003** GreyNet relaunch; GL5 – Fifth International Conference on Grey Literature in Amsterdam with INIST as Co-Sponsor and EAGLE on Program Committee **2000** GreyNet discontinued	**1999** GL'99 – Fourth International Conference on Grey Literature in Washington D.C. USA **1997** GL'97 – Third International Conference on Grey Literature in Luxembourg with EC as Host and EAGLE on Program Committee **1995** GL'95 – Second International Conference on Grey Literature in Washington D.C. USA; **1993** GL'93 – First International Conference on Grey Literature in Amsterdam with EAGLE as Sponsor and on the Bi-Annual GL-Program Committees 1993-1999 **1992** GreyNet founded	**1985** EAGLE Association was founded as producer of the SIGLE Database **1980** SIGLE established as database and document delivery system; Later exploited via STN and Blaise Hosts as well as SilverPlatter CD-ROM **1978** York Seminar on Grey Literature hosted by British Library and the EC

Part II, Section Four

Applications and Uses of Grey Literature

What do we know about the usage and impact of grey literature? Gentil-Beccot reports from recent studies pertaining to the use of information in the High-Energy Physics (HEP) community, where survey data and citation as well as log analysis is employed. Her chapter contains interesting details and observations, for example on the link between published articles, preprints and other grey literature, as well as on peer-review, digital libraries, and the role of the scientific community. She provides evidence of the growing importance of grey literature for communication in the new technological environment relevant to her discipline. She further reveals that today HEP scientists ask yet for even more and see "access to data and tables as important, (…) another essential aspect of the future of GL in high-energy physics".

The second chapter in this section assesses the real use and impact of grey literature by public institutions. "(…) thousands of studies are conducted, and tens of thousands of print and digital reports are produced annually, many of which have direct or indirect policy implications. What is poorly documented is whether adequate attention is paid to such reports, which are typically grey literature, and to subsequent advice, both by sponsoring agencies and by other users." Here, MacDonald [et al.] examine empirical results from ongoing research (citation and survey data) and conclude with ten recommendations to improve awareness, retrieval, use, and the standing of grey literature.

While data from MacDonald [et al.] emanate from the Marine Sciences, the third chapter in this section draws upon survey results from the Geosciences, namely karst research. Chavez confirms that grey literature is regularly used but less frequently cited. In so doing, he confronts the limited use of Web2.0 tools on geoscience platforms. Here again, the interaction between content and IT environment, e.g. digital library, infrastructure, content management and added value services become manifest. While content is king, content needs environment. In his concluding remarks, Chavez emphasizes the "value of a library-led collaboration with (…) research communities".

The final chapter in this section investigates a specific sector of the information market by examining the use of grey literature produced by non-governmental organizations (NGOs). The research by Crowe [et al.] focuses on healthcare information in developing countries. "NGOs create grey literature in the form of reports, online newsletters, blogs, etc. However, (…) there is a need to increase involvement of NGOs in the management of their knowledge output." The authors

argue in favour of partnerships with information services and other such agencies in the implementation of dedicated open repositories. And, their chapter concludes with a model or framework meant to improve preservation and dissemination of grey items.

Based on the work of the authors in this section, we can draw upon consensus that a lot still remains to be done. Today we have considerable knowledge regarding the usage of digital online resources such as journals, articles, databases, and e-books. However, much less is known on the usage and impact of grey items, especially in open archives. While the standardization of metrics and tools is ongoing, we nevertheless need more usage data - especially from surveys and deep log analyses.

For a better understanding of this, the reader would do well to consider the following three lines of questioning: What kind of (basic) empirical data and metrics do we need to assess and compare the usage of grey items? How can we assess impact and usage in different environments e.g. scientific communities as opposed to political communities? And ultimately, how can we best describe the link between the IT environment and usage?

Chapter 10

The driving and evolving Role of Grey Literature in High-Energy Physics

Anne Gentil-Beccot, CERN, Switzerland

10.1 Introduction

If grey literature (GL) is often seen as a marginal part of the scientific information landscape, this is absolutely not true for high-energy physicists (HEP) who developed, decades ago, their own scientific communication scheme using this allegedly "darker" fraction of the literature. Today, grey literature remains a living and indispensable resource for this discipline. What is more, grey literature has become a driving force, motivating many evolutions in the HEP information landscape.

At the same time, publication in journals continues to be essential for the scientific community, most of the preprints being eventually published. But new challenges in scientific publication, such as Open Access publishing, are under discussion nowadays. Furthermore, the information landscape is becoming increasingly complex. In addition to the tools developed by the community, scientists can use many information products such as commercial databases or search engines like Google or Google Scholar - users can now access, easily or not, huge amounts of varied information. In addition, needs are changing, new technologies appear every day and new ways of interacting with users evolve. In such a context, it is interesting to analyse the current role of grey literature in HEP, and try to understand what the future evolution might be.

After describing in detail the pervasive role GL has taken in HEP for several decades and how the community has developed dedicated tools adapted to these specific resources, we will show that HEP scholars continue to rely on this literature for accessing information because it meets their need for fast communication. Finally, we will show how the community is continuously adapting its information tools, evolving with the needs of the users and the fast-changing technologies.

10.2 The pervasive “preprint culture”

High-energy physics aims at discovering the constituents of matter and understanding their interactions [1]. This is a small and cohesive community, counting around 30,000 scientists, with a strong collaborative spirit. Because of its special characteristics, this community has developed its own scientific communication scheme, mostly based on grey literature.

10.2.1 Some historical perspectives

Fifty years ago, long before the birth of the online world, the delay between the submission of a scientific paper and the time it reached the reader appeared unacceptable to HEP scientists, who therefore adopted preprints as their main communication channel. The community was already composed of two sub-groups who needed to communicate both internally and externally: experimental physicists working at accelerators of ever-increasing energy, regularly witnessing new discoveries during the early stages of the discipline; and theoretical physicists interpreting these results, improving their theories and suggesting new projects. It was simply out of the question to accept months of delay in communication - the average turnaround time of ideas and experimental results was no more than several months - grey literature was the solution!

For decades, theoretical physicists and experimental collaborations, wanting to disseminate their results in a way faster than the distribution of conventional scholarly publications, printed and mailed to all major HEP institutes copies of their manuscripts at the same time as submitting them to peer-reviewed journals [2,3,4]. The same institutes financially supported the dissemination of the scientific results of their researchers, and this implied high costs[1]. Libraries spent also resources indexing all these preprints, working papers and reports, making them accessible to the institute’s researchers.

In a sense, this “preprint culture” in high-energy physics pioneered the “open access” distribution of scientific results. This form of “institute-pays” Open Access ensured the fastest and broadest possible dissemination of scientific results. It is worth noting that this process favoured scientists working in well off institutions. These could pay for the mass mailing and were most likely to receive copies of preprints from other scientists seeking recognition. Smaller and less well off institutions had therefore less chance to disseminate their results and become aware of the research of other scientists.

1 In 1990, CERN used to spend over 1 Million Swiss francs a year for printing and mailing expenses.

10.2.2 Development of dedicated tools

With the increasing use of the Internet, the process continued electronically and the cards could be redistributed more equally. Indeed, the community launched its own tools to manage grey literature: in 1991, even before the web was invented, Paul Ginsparg, at the Los Alamos laboratory, launched arXiv [5], the first physics preprint repository. This new tool ensured the transition from an old preprint world to a new electronic system, offering all scientists an easy and less-restricted way to access and disseminate information, by removing the cost-barrier of mass-mailing preprints all over the world.

With more than 500,000 articles, arXiv has today grown beyond the field of HEP, becoming the major repository for many other disciplines [6].

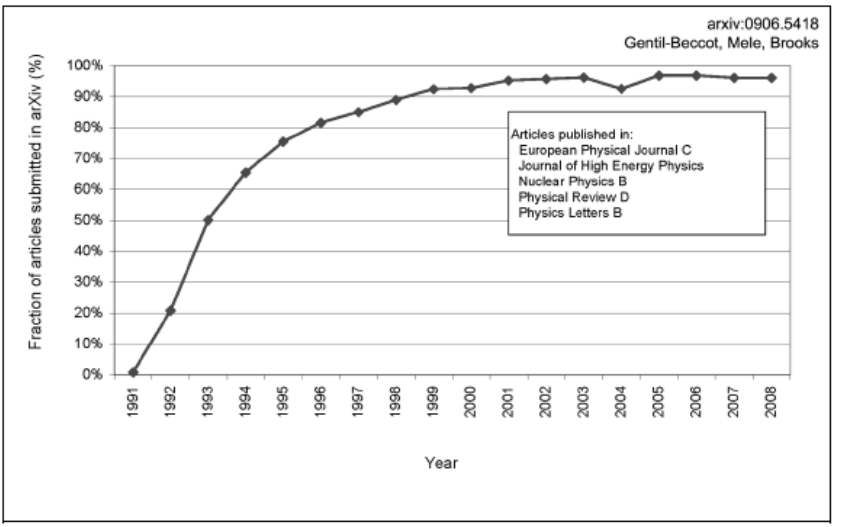

Figure 1. Fraction of articles published in the main peer-reviewed HEP journals which also appeared, in some version, on arXiv.org.

The SPIRES database [7], the first grey literature electronic catalogue [8,9], was born at SLAC (Stanford Linear Accelerator Center) HEP laboratory in Stanford, California, in 1974, and was developed in collaboration with DESY, in Hamburg, Germany and Fermilab, Chicago. It listed preprints, reports, journal articles, theses, conference talks and books, and it now contains metadata for about 760,000 HEP articles. This tool took advantage of the invention of the web, and became the first web server [10] in the U.S. In summer 1992, SPIRES linked to the arXiv for full-texts, starting a close partnership, and bringing preprints onto the web, accessible through detailed indexing including reference to the published versions, when available.

The community produces around 5,000 journal articles per year. The large majority of these articles are published in just six peer-reviewed journals from

four publishers [11]. In figure 1, we see that 90 to 100% of the articles published in these six journals are also submitted to arXiv; we see also that this situation is stable and has lasted for ten years already. In addition, it is important to say that many HEP scientists routinely upload to arXiv a revised version of their preprint, which matches the final peer-reviewed version, including the corrections introduced during the publication process.

Even in the era of electronic journals, grey literature fully retains its importance in the discipline; arXiv today contains the vast majority of preprints (in most of the cases in their peer-reviewed version) in the field, this means that almost the entire literature in the discipline is freely accessible on the web. It is important to say this situation happened without any debate or mandate, driven by the specific needs of the scientists, as we shall see in the next section.

10.3 GL enabling fast and immediate communication

Almost the whole literature produced in HEP is available on arXiv. But why is this a reality in high-energy physics, while in many other disciplines repositories are hardly filled?

10.3.1 A direct benefit for the community

A study [12] has been carried out in 2009 on the actual usage of information in the community using citations and logs analysis.

We see in Figure 2 one of the findings of this study. Articles published in two leading HEP[2] journals over 10 years were split into two samples, those which were submitted to arXiv (96.4% of the total) and those which were published without appearing on arXiv (3.6% of the total); we see clearly that articles submitted to arXiv begin accumulating citations long before publication. This shows that in HEP the scientific discourse happens when the literature is in its "grey" stage. Citation begins well before publication, because authors read the preprint earlier. It is worth adding that, in this graph, citations from preprints have been taken into account, this explains the peak of the top line appearing almost at the publication date, since no publication delays (either from the citing paper, or from the cited paper) are taken into account, and this also explains the very steep rise of the bottom line, as here again the citing paper, which might be in a preprint form, is citing immediately after publication of the cited paper. This demonstrates twice that scientific discussion starts much before publication in a journal.

Hence, there is an immense incentive for scientists to use grey literature: the speed of information dissemination. And, in the same way that in the 1960s scien-

2 Physical Review D (published by APS) and Journal of High Energy Physics (published by IOP/SISSA till December 2009 and by Springer/SISSA as from January 2010)

tists mass-mailed preprints to disseminate their research results as fast as possible, today, they use arXiv for the very same reason, with, obviously, much greater efficiency.

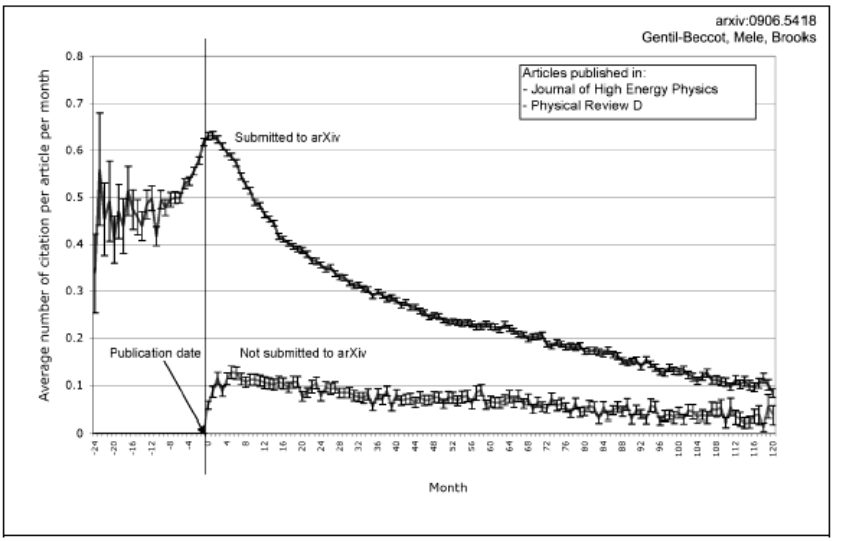

Figure 2. Average number of citations per article per month as a function of the time of the citation relative to the time of publication. Data is from 26,741 articles from the Journal of High Energy Physics and Physical Review D over the period from 1998 to 2007.

While the above citation analysis gives an understanding of the speed and manner of the scientific discourse in HEP, the analysis of click streams provided by SPIRES gives even more information about the actual reading habits of HEP scientists.

10.3.2 The predominance of grey literature even for peer-reviewed information

A survey [13] performed in 2007[3] demonstrated that about 50% of HEP scientists use SPIRES for a bibliographic search. Therefore, the analysis of click streams of SPIRES users once an article has been identified and can be accessed, gives a

3 From April 30 to June 11 2007, a survey was jointly conducted by several High Energy Physics institutes: CERN (European Organization for Nuclear Research), SLAC (Stanford Linear Accelerator Center), DESY (German Electron Synchrotron) and Fermilab (Fermi National Accelerator Laboratory). The aims of the survey were to understand the users' perceptions of current HEP information systems, to assess user requirements and preferences, and to define future needs. During the survey period, 2110 answers were received.

clear representation of the reading habits of the community. SPIRES click streams collected during October 2008 have been analysed (30,000 clicks) [12]. The study was restricted to clicks that occurred from records displaying both a link to arXiv and to a publisher website. We discovered that in 82% of cases, arXiv is preferred. The survey showed also that 40% of the scientists go directly to arXiv to access information, they are therefore not counted in the data mentioned here. The advantage of arXiv over the published version might thus be much higher than what is shown by the click streams analysis.

This brings us to the fact that HEP scientists prefer to read the arXiv version of published papers, giving grey literature an even more important place. Several typical characteristics of the community help to explain this result. One of the main reasons is that, in most of the cases, the author resubmits a revised version of the preprint including the corrections brought by the peer-review process. Besides, arXiv provides free access to its content, whereas the published version on the journal website is often under subscription, restricting access. It confirms a result from the 2007 survey: respondents were asked which system they used the most when looking for preprints or published articles in different search situations. It became clear that the overall landscape does not change substantially between preprint and article searching. This is extremely relevant in the interplay between grey literature and "conventional" literature in the field: when HEP scholars need to access "conventional" literature they still use the systems that were initially conceived to index and curate the grey literature in the field! Solutions invented for grey literature are therefore mainstream in this community.

10.3.3 A redistribution of the roles in the HEP information landscape

As demonstrated, peer-reviewed journals have lost their role as providers of information and as a means of scientific communication, which has effectively moved to the grey literature. However, HEP peer-reviewed journals continue to play an indispensable role, providing independent quality control, which is necessary in this field as in the entire academic community. This situation, far from increasing the gap between GL and published literature, allows the clarification of their mutual role which leads to their separate but synergic evolution. There is no space here for further discussion of this issue, but the SCOAP3 project [14,15] aims to convert all HEP journals to Open Access, according to a model where the peer-review role of journals, rather than the dissemination, would be financed by the community.

In addition, the success of GL in HEP is due to the fact that GL in this field has taken advantage of the new technologies, adapting and shaping them, rather than resisting and retrenching. This is the topic of the next section.

10.4 Evolving needs, evolving tools

The survey performed in 2007 [13], mentioned above, aimed to assess the current usage of information resources by HEP scholars as well as their future needs. This section will discuss more specifically some of these results.

10.4.1 The predominance of "community-made" tools: today and tomorrow

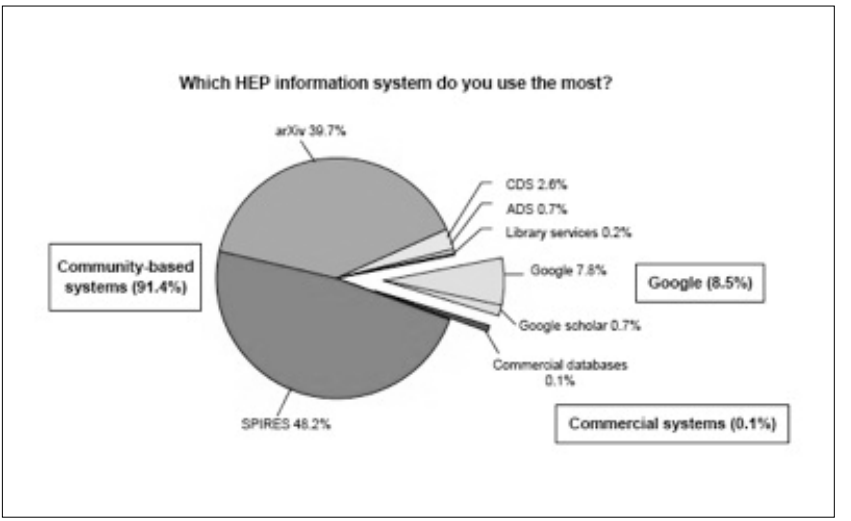

Figure 3: Favourite information resources for HEP scholars.

Thanks to the association of arXiv and SPIRES, scientists have access to the whole HEP literature, either in the preprint or the published version, SPIRES providing detailed metadata and publication information, and arXiv providing full-text preprint versions of nearly all journal articles. As we mentioned in the previous section and as shown in figure 3, the survey demonstrated that arXiv and SPIRES were the two main tools used by the community. 91.4% of the respondents prefer community-based services[4]. This is not a surprising result since most of these community-based systems were created to meet the needs of HEP scholars, and these tools tend to be tailored specially to the evolving needs of the HEP community. These systems have been user-driven for decades.

On the other hand, 9% of the respondents claimed to use Google or Google Scholar. The survey also showed that the use of Google tends to increase as the

4 CDS (CERN Document Server)[16] and NASA ADS [17] are also tools created by the community.

age of the respondents decreases. This is a reflection of the integration into the HEP community of scholars belonging to an age group that has been exposed to Internet search before their academic career, as opposed to scholars who first used Internet search engines during their professional activity. It must be noted that Google might actually be only a gateway to other sources, as it indexes material from arXiv and SPIRES. Effectively, those using Google ultimately use some of the other information services to access the document. This result cannot therefore be seen as proof that Google is becoming more important than the community tools, it only shows the increasing need for users to have a single entry to their information resources.

In summary, the survey demonstrated that community-based systems largely dominate the landscape, even if Google takes a non-negligible part. The choice of these information systems corresponds to the need of scholars for easy access to full-text and a wide coverage of all literature in HEP, which is exactly what is currently offered by the combination arXiv.org & SPIRES.

10.4.2 Access to even more grey literature

The survey, however, showed a need from a large part of the users to have access to emerging forms of grey literature such as conference slides. Scientists go to conferences to present and discuss the latest research results. Slides shown during conference talks generated in digital format, constitute a new form of grey literature that other scientists want to access immediately, and often quote in their subsequent publications, without having to wait for a conference paper to be written and submitted as a preprint to a repository [18]. This is indeed the next frontier: capturing, storing and indexing the content. It is important since this information is not always organised: links to conference slides might get broken, as might the web sites of conferences. Thus, there is a need for a system to harvest and serve the content. Some projects aim to organize this information such as Indico [19], but not all conferences benefit from such tools and this is a new call to action for information providers. In addition, the survey underlined the need for a single interface to access all required information, whatever its type.

Another important field of improvement requested by the users is access to theses. Regarding theses, the survey shows that users access Google much more than for any other information type. This emphasizes that community-based systems do not yet cover the complete scientific information needs of the field, particularly in the area of grey literature represented by theses. While the vast majority of HEP literature can be found on arXiv and in a few major journals, conference proceedings and theses are distributed over a multitude of servers and are thus more difficult to collect. Also, commercial databases do not provide a better service in this regard. On the contrary, only Google has an advantage here since it indexes a lot of resources.

By way of conclusion, this then suggests further development of the community-based services with investment in the harvesting and preservation of theses as well as other non-peer-reviewed user-generated content.

This is once again proof of the growing importance of grey literature for the immediacy of communication in the discipline. Furthermore, many users see access to data and tables as important, yet another aspect of the future of GL in high-energy physics. Indeed, data such as tables behind graphs can also be considered a form of GL, since their publication is not yet standardised. The way these data are (will be) described, preserved and accessed constitutes the next frontier for community information systems, since they have the advantage of being close to the community, while publishers are only starting to think of how to provide these data for their users.

10.4.3 Integration of Web 2.0 technologies

Future needs imagined by users correspond primarily to a wish to have easier and wider access to content. But in a context where our daily communication channels include more and more Web 2.0 technologies, other needs appear, such as 'recommendation of articles' (almost 50% of the respondents think it is somewhat important).

Furthermore, a question in the survey tried to assess the potential for the implementation of Web 2.0 features to capture user-tagged content. We find that 63% of the respondents claimed they would be willing to spend between five minutes a day and an hour a week, showing that there is immense potential for user-tagged and user-curated content in the field of information provision in HEP. This question is essential because users need to retrieve correct and accurate information. This user-tagging and user-curating could help future information systems to provide accurate information to their users. We don’t know yet how this will evolve, how far the community will use these new tools. But, this type of information, which can be considered as unstructured, will without any doubt be scientific content that will be put online and will have a value as such.

To summarize, many challenges appear when one starts analysing the wishes of the users, but they outline the best track to follow in order to build a system fully adapted to the community usage.

10.4.4 Towards a new system

The survey and discussions between the four leading HEP laboratories (CERN, DESY, Fermilab and SLAC), in synergy with other partners (notably arXiv) and in a continuous dialogue with major publishers in the field, led to the idea of the next generation HEP information system, merging the current SPIRES database and a modern platform, the Invenio open-source digital library software [20]. This

new information system, INSPIRE [21], is being developed by a collaboration of the four leading HEP laboratories mentioned above. It will integrate the content of present repositories and databases to host the entire set of metadata and the full-text of all open access publications, past and future, including conference material, becoming the single entry for the whole community to all HEP relevant material.

Grey literature has for decades been the driving information source for HEP scientists and will maintain this role even through the future evolutions of information systems.

10.5 Conclusions

With this last section, we conclude our snapshot of the HEP community scientific information practices. We saw that the community has always been anticipating and driving the evolutions. Grey literature became the most important communication channel because the community needed immediate access to information. This is still the case, and this is why the community develops new systems that make the access to information even more immediate. But needs evolve and the community must now go further: open access to full text is not enough anymore, scientists want improved access to greyer literature, such as conference proceedings, theses, or high-level data, not yet available anywhere. They need to interact more deeply with all these resources. This is one of the challenges the new HEP information system INSPIRE will meet.

Another problem that will have to be addressed is the changing role of published literature, which no longer serves any communication purpose, even if its place is still vital for other scientific reasons.

But one constant remains in all these movements, the community itself takes the lead in all these changes, and that's why it is well placed to successfully achieve these evolutions, as they are driven by no more than its own needs.

10.6 References

[1] CERN, public webpages
http://public.web.cern.ch/public/en/Science/Science-en.html
[Last visited October 06, 2009]
[2] Goldschmidt-Clermont L. (2002), Communication Patterns in High-Energy
Physics, in High Energy Physics Libraries Webzine, 6 (2002)
http://library.web.cern.ch/library/Webzine/6/papers/1/
[Last visited October 06, 2009]
[3] Heuer, R.D., Holtkamp, A., Mele S. (2008), Innovation in Scholarly
Communication: Vision and Projects from High-Energy Physics, in Information Services & Use 28 (2008) 83 – 96, arXiv:0805.2739.

[4] Aymar R. (2009), Scholarly communication in High-Energy Physics : past, present and future innovations, in European Review 17 (2009) 33 – 51, CERN-OPEN-2008-015.

[5] ArXiv website

http://www.arXiv.org

[Last visited October 06, 2009]

[6] Ginsparg, P. (1994), First Steps Towards Electronic Research Communication, in Computers in Physics 8 (1994) 390.

[7] SPIRES Website

http://www.slac.stanford.edu/spires

[Last visited October 06, 2009]

[8] Addis, L. (1994), Brief and Biased History of Preprint and Database Activities at the SLAC Library, 1962-1994

http://www.slac.stanford.edu/spires/papers/history.html

[Last visited October 06, 2009]

[9] Kreitz, P.A. and Brooks T.C. (2003), Subject Access through Community Partnerships: A Case Study, in Science & Technology Libraries, 24 (2003) 153, arXiv:physics/0309027

[10] Kunz, P. et al, The Early World Wide Web at SLAC

http://www.slac.stanford.edu/history/earlyweb/history.shtml

[Last visited October 06, 2009]

[11] Mele, S, Dallman, D, Vigen, J., Yeomans, J. (2006), Quantitative Analysis of the Publishing Landscape in High-Energy Physics, in Journal of High Energy Physics 12 (2006) S01, arXiv:cs.DL/0611130

[12] Gentil-Beccot, A., Mele, S. and Brooks, T.C. (2009), Citing and Reading Behaviours in High-Energy Physics. How a Community Stopped Worrying about Journals and Learned to Love Repositories

arXiv:0906.5418

[13] Gentil-Beccot, A., Mele, S., Holtkamp, A., O'Connell, H. and Brooks, T.C. (2009), Information Resources in High-Energy Physics: Surveying the Present Landscape and Charting the Future Course, in Journal of the American Society for Information Science and Technology, 60 (2009) 150–160, 12

arXiv:cs.DL/0804.2701v2

[14] Bianco, S. et al., Report of the SCOAP3 Working Party, ISBN 978-92-9083-292-8

http://scoap3.org/files/Scoap3WPReport.pdf

[Last visited October 06, 2009]

[15] SCOAP3 website

http://www.scoap3.org/

[Last visited October 06, 2009]

[16] CERN Document Server

http://cdsweb.cern.ch

[Last visited October 06, 2009]

[17] NASA Astrophysical Data System

http://adswww.harvard.edu/

[Last visited October 06, 2009]
[18] Gentil-Beccot, A. (2008), How do High-Energy Physics scholars discover scientific information? In TGJ, 4 (2008).
[19] Indico website
http://cdsware.cern.ch/indico/
[Last visited October 06, 2009]
[20] CDS Invenio website
http://cdsware.cern.ch/invenio/index.html
[Last visited October 06, 2009]
[21] Inspire website
http://hep-inspire.net/
[Last visited October 06, 2009]

Chapter 11

The Use and Influence of Information Produced as Grey Literature by International, Intergovernmental Marine Organizations: Overview of Current Research

Bertrum H. MacDonald, Peter G. Wells, Ruth E. Cordes, Gregory R. G. Hutton, Danielle M. Cossarini, and Soonai S. Soomai
Dalhousie University, Canada

11.1 Introduction

The great abundance of data and information in most fields of enquiry is a dominant characteristic of our time (circa 2009). One estimate suggests that 988 billion gigabytes of information of all kinds will be generated in 2010, compared to five billion gigabytes created only six years earlier (Palfrey and Glasser, 2008). The increased rate of producing new information is by itself astonishing but the ease by which publication can now occur, especially as "grey literature" in all of its forms, including open access, is a matter of growing interest.

An emerging view from the marine environmental field is that the large storehouse of available information, much of it in the grey literature, needs to be more effectively used to solve urgent global issues (Thatje, Laudien, Heilmayer and Nauen, 2007; Wells, 2003). For example, problems of awareness persist, even though most of the new information is now digitally produced and arguably easier to access. It is now recognized that the diffusion, use, and influence of such information are complex and variable processes (de Alwis, Majid, and Chaudhry, 2006; Evans and Reimer, 2009; Healy and Ascher, 1995; Holmes and Clark, 2008; McNie, 2007), and given the problems to resolve, they are a priority for investigation.

Governmental and intergovernmental bodies, long known as prolific writers and frequent publishers, contribute to the growing body of information. Since political, economic, and environmental issues frequently transcend regional and national borders, these bodies have often been set up to play significant roles in seeking solutions to today's serious global environmental problems. Hence, thousands of studies are conducted, and tens of thousands of print and digital reports are produced annually, many of which have direct or indirect policy implications.

What is poorly documented is whether adequate attention is paid to such reports, which are typically grey literature, and to subsequent advice, both by sponsoring agencies and by other users. Such documentation is needed for accountability and tracking progress on problem resolution.

Our study of intergovernmental organizations, begun in 2001, is focussing on marine environmental and fisheries information. We are learning how such organizations produce, publish, and disseminate grey literature, and how they promote awareness, access, and use of it. Our goal is to understand how pertinent information produced by these bodies can be more effectively used in decision making processes. In this chapter, we present an overview of our ongoing research on the scientific grey literature of three intergovernmental bodies: the UN-based Joint Group of Experts on the Scientific Aspects of Marine Environmental Protection (GESAMP), the Gulf of Maine Council on the Marine Environment (GOMC), and a Regional Fishery Body of the UN Food and Agriculture Organization (FAO). Insights gained from this research are described, and recommendations are provided to improve the use and influence of this information.

The Intergovernmental Panel on Climate Change (IPCC) provides an example as to why it is important to address questions of communication and use of information in decision making. For over two decades, the IPCC has drawn worldwide attention to "the importance of climate change as a topic deserving a political platform among countries [in order] to tackle its consequences" (Bolin, 2007; IPCC, n.d.). As one of many intergovernmental scientific bodies, the IPCC has probably had more success than most in communicating its information, primarily as grey literature (technical reports and policy summaries), and influencing public policy on climate change at local, national, and international levels. However, a recent IPCC report concluded that "communication of complex scientific issues remains a difficult task" (IPCC, 2009, p. 2). If communication continues to be a "difficult task" for an organization of IPCC's stature, do other bodies that produce significant publications and advice also experience a similar challenge?

Two additional questions may also be asked: 1) Has the profile and use of grey literature produced by governmental and intergovernmental bodies increased with the transition from solely print to a digital publication universe? and 2) Has the recent flood of information resulted in a greater challenge for placing timely and salient information on the agendas of policy and decision makers when they need it? We are reflecting on these questions as the research continues.

11.2 Study Framework, Methodology, and Case Studies

Determining the paths that scientific publications take and developing an understanding of the use and influence of their information content are not trivial tasks. Neither task is linear and both are likely subject to serendipity and unknown influences. However, we believe that an approach employing various information research methodologies, including citation analysis, document content analysis,

online surveys, and interviews, can lead to an appreciable increase in understanding the use and influence of grey literature. As well, the problem of how to better utilize existing information usually does not lie with the lack of information, but with its communication. Hence, our research focuses on the interface between production of scientific grey literature and its use primarily in policy and decision-making contexts (see Figure 1). Using this guiding framework, we are developing techniques to measure use and influence, and to identify and mitigate communication barriers.

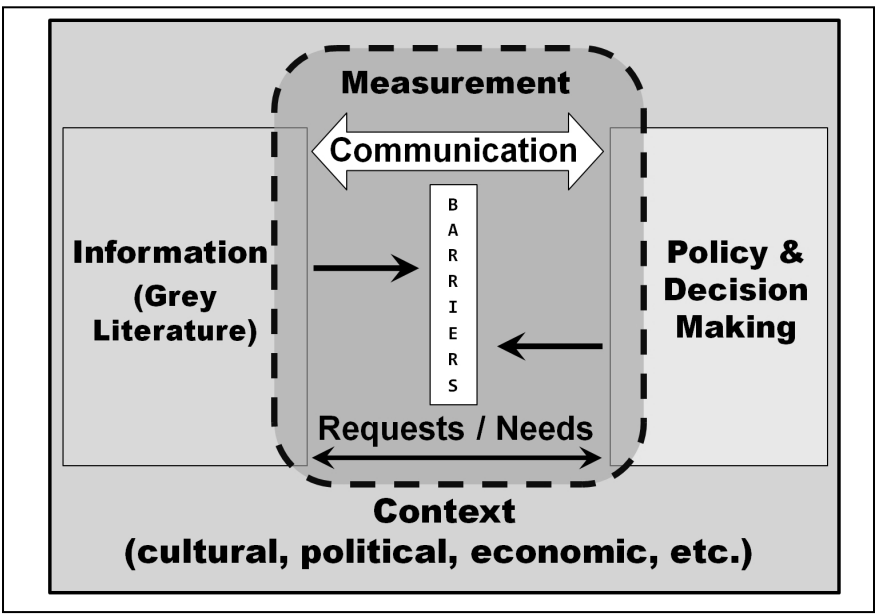

Figure 1: Study Framework: Use and Influence of Grey Literature

Various methods are being employed while analysing the grey literature of the intergovernmental groups, including identification of the organizations' publications, analysis of citations to those publications, and surveys of key informants (Cordes, 2004; Cordes, MacDonald, and Wells, 2006; Hutton, 2009; MacDonald, Cordes, and Wells, 2004, 2007; Soomai, 2009).

11.2.1 Case 1: The Joint Group of Experts on the Scientific Aspects of Marine Environmental Protection (GESAMP)

GESAMP is a leading United Nations scientific advisory body on marine pollution and marine environmental protection. It has been producing significant reports for almost forty years on marine pollution and the protection and management of marine living resources and ecosystems, especially where

multidisciplinary advice proves beneficial (Pravdić, 1981; Wells, Duce and Huber, 2002; Windom, 1991). The reports are published by the sponsoring UN agencies. At GESAMP's annual sessions, scientific members (appointed by the GESAMP secretariat, which is made up of representatives of the sponsoring UN agencies) review their work program, receive reports, approve publication of reports after thorough review, and discuss emerging issues affecting the oceans. Most of the substantive work on specific issues, presented annually, is carried out intersessionally by designated working groups under agency sponsorship (e.g., Wells, Höfer, and Nauke, 1999). Chaired by a GESAMP member, each working group includes invited marine specialists from around the world. Meetings of the working groups are highly technical, with the goal of producing detailed reports on specific topics (e.g., oil pollution and invasive species). Most groups are assigned specific tasks that can be accomplished in one to three years, and disband after their reports are reviewed (internally and externally), revised, and published in the GESAMP *Reports and Studies* series. Some groups have had lengthy histories, however, and have produced many reports. For example, the Working Group on the Evaluation of the Hazards of Harmful Substances Carried by Ships (EHS) began as an *ad hoc* panel in 1972, but since 1974 has had a major role in evaluating the hazards of chemicals carried by ships for the International Maritime Organization and the MARPOL 73/78 Convention (Wells, Höfer, and Nauke, 1999).

GESAMP's sponsoring agencies have published 77 GESAMP reports to date (GESAMP, 2008), including major periodic health of the ocean assessments, such as Report # 39, *The State of the Marine Environment*. Although many scientists consider "grey literature" or technical reports generally not peer reviewed (e.g., Natural Resources Canada, 2007), that is not the case for GESAMP's reports. Early reports did not explicitly acknowledge the thorough, open reviewing process, but reviewing details have been included recently. For example, Report # 64, *Hazard Evaluation Procedure of Chemical Substances Carried by Ships*, was the result of six years of work by the thirteen-member EHS Working Group and was refereed by nine external experts, Report # 70, *A Sea of Troubles*, lists over 90 individuals with various roles in its preparation, and Report # 75 on oil pollution had over a dozen independent technical reviewers.

11.2.2 Case 2: The Gulf of Maine Council on the Marine Environment (GOMC)

Formally created by the Premiers of the Canadian provinces of Nova Scotia and New Brunswick and the Governors of the American states of Maine, New Hampshire, and Massachusetts in December, 1989, the Gulf of Maine Council on the Marine Environment (GOMC) focuses on the marine environment of the Gulf of Maine and the Bay of Fundy, in the North-west Atlantic Ocean (MacDonald, Cordes, and Wells, 2007). GOMC is a bilateral intergovernmental body, with linkages to non-governmental organizations (NGOs) and the university research

sector. It pays particular attention to marine environmental issues and their resolution, particularly those of a cross boundary nature (e.g., air and water pollution, conservation of critical habitats and hemispheric migratory species, climate change, and introduced species). Overall, the GOMC's work entails research, ecosystem monitoring, communication and education, and public policy.

The Council develops and works under the guidance of five-year action plans. The current action plan has three primary goals: habitat conservation and restoration, human and ecosystem health, and environmental sustainability (GOMC, 2007). Among GOMC's several long-term initiatives are the Gulfwatch contaminants monitoring program, a salt-marsh restoration program, its quarterly newspaper (*The Gulf of Maine Times)*, and the Council's Website (www.gulfofmaine.org). The Council Secretariat rotates among the five states and provinces on an annual basis, and is chaired by an individual in the host jurisdiction. The Council's mandate is carried out primarily through its core Working Group, which reports to the Council, and is also chaired by a representative of the host jurisdiction on a one year cycle. Several committees and subcommittees, co-chaired by American and Canadian members, meet at least once per year and report to this Working Group. The core work of GOMC is conducted by individual researchers and through the work plans of the member agencies. The Council's significant communication agenda is pursued primarily through its Website, its publications (now placed onto the Website), and many workshops on a variety of topics.

11.2.3 Case 3: The Food and Agriculture Organization (FAO)

The FAO provides financial and technical support for a number of global and regional projects related to fisheries management and sustainable development. The Western Central Atlantic Fishery Commission (WECAFC), a regional fishery body, was established in 1973 under Article VI-1 of the FAO Constitution to address the marine conservation and development needs by Member Nations or countries interested in the fisheries of the management area (FAO, 2009). This marine area extends from Cape Hatteras, United States (35 degrees N), to just south of Cape Recife, Brazil (10 degrees S), hence covering the southeast coast of the United States, the Gulf of Mexico, the Caribbean Sea, and the northeast coast of South America. The area is politically and geographically complex, with an equally complex marine biodiversity (Soomai, 2009). WECAFC has an advisory management function and facilitates research, education and training, and assists its members in establishing policies to promote the joint regional management of resources.

WECAFC has established a number of subsidiary bodies (Working Parties) in which much of the advisory work is done, and *ad hoc* working groups which conduct research and assessment of marine resources within defined geographic areas. One such group is the Working Group on Shrimp and Groundfish Resources in the Brazil-Guianas Shelf. The governments of countries of the Brazil-Guianas Shelf

region ─ Trinidad and Tobago, Venezuela, Guyana, Suriname, French Guiana, and Brazil ─ interact with the FAO in implementing programmes and meeting their responsibilities for national and regional assessment and management initiatives under the FAO/WECAFC Shrimp and Groundfish Working Group. This Working Group is made up of the technical staff of the FAO, technical consultants, and national scientists. Industry representatives, members of NGOs, fisheries managers, and policy makers are invited to attend periodic Working Group meetings (Soomai, 2009).

Since its inception in the 1970s, the WECAFC has coordinated annual technical meetings to conduct scientific assessments of the shrimp and groundfish resources of the region. It has produced numerous, grey literature technical reports that are available on the FAO Web site. Print copies are distributed to fisheries departments and research institutes of each WECAFC member country (FAO, 2009).

11. 3 Results and Discussion

Determining the use and influence of grey literature publications begins with identifying the processes by which organizations prepare and produce such publications and with describing the total published output of those organizations. These tasks are not always straightforward.

Our case studies show a variety of methods and types of publication. GESAMP's methods and publications (working documents, technical reports, journal papers) are quite well understood and documented (Cordes, 2004; Hutton, 2009; MacDonald, Cordes, and Wells, 2004; Wells, Duce, and Huber, 2002). In contrast, the GOMC, working by itself and in collaboration with others, has produced a wider variety of publications, including conference proceedings, technical reports, conference background documents, annual reports, action plans, newsletters, newspapers, magazines, fact sheets, brochures, maps in poster format, and videos (Cordes, MacDonald, and Wells, 2006). Moreover, individuals associated with the Council have given many workshop and conference presentations and written primary journal articles. Confirming this diverse array of publications was complicated by the absence of a central repository holding the output of the Council. As well, unlike for-profit or commercial publishers which systematically apply publication standards and aggressively promote their publications, intergovernmental organizations often overlook the benefits of such standards and practices. Hence, the output of an organization may be poorly or incompletely documented, possibly diminishing the use and value of the group's published information.

Verifying GOMC's publications required locating evidence from a variety of sources, including GOMC's annual reports and Website, library catalogues, Web search engines, article databases, electronic collections, and conference proceedings. The searches also led to the discovery of items published by other organizations with GOMC's support, and evidence of other publications that had effec-

tively vanished, as electronic versions are no longer on the Web and print copies are not in library holdings. This process of discovery showed that a current and complete inventory, and tracking method, for a group's publications is essential for ensuring their widest awareness, use, and possible influence.

While intergovernmental organizations have an interest in determining the use and influence of their publications, how use and influence is measured has received limited attention. Some studies of scientific publications have been based on citation data, primarily available through the Science Citation Index, now Web of Science, since the mid-1950s (Bar-Ilan, 2008). Although competitors to Web of Science, e.g., Elsevier's Scopus, have become available more recently, neither Web of Science nor Scopus provide extensive coverage of grey literature. They were not designed to record and track grey literature in all of its breadth. The limited citation data coverage for grey literature in Web of Science and Scopus is complicated by increasingly varied forms of online scientific publications (Borgman, 2007; Vaughan and Shaw, 2005). This situation clearly shows that Web of Science or Scopus cannot be relied upon, in their present form, for documenting the use of grey literature produced by the groups under study. A similar finding was made by researchers who recently examined the publications of the North Pacific Marine Science Organization (PICES) (Voss and Webster, 2007).

Continued study of citation data for GESAMP publications showed that a composite metric of use and influence can be developed by analysing citation data obtained from several sources: Google, Google Scholar, Web of Science, and monographs (Hutton, 2009). Inclusion of citations in monographs, reports, and other Web-based materials allows a more complete understanding of use and influence of this grey literature.

However, as informative as citation data are, such evidence misses measuring the use and influence of information in publications in contexts where citing other work is uncommon or does not occur, or where documentation is proprietary or not generally in the public domain, e.g., ministerial briefing documents, strategic planning papers, and action plans, etc. Specifically, citation studies do not completely document information use and influence in public policy and decision making contexts. But in these domains, as long as material is accessible, content analysis of internal documentation can provide insights regarding how published information may have been consulted, debated, and applied. Further understanding of a document's influence can be obtained through interviews with key informants, as our on-going studies have begun to show (Cossarini, 2009; Soomai, 2009; Wells, MacDonald, Cordes, Hutton, Cossarini, and Woods, 2009).

Intergovernmental bodies often see their primary responsibility as offering solutions to problems through the production of expert information and reports. They are rarely able to implement communication strategies, especially when a dissemination role and appropriate personnel are absent in the organization. In our research, this was the case with GOMC and FAO-WECAFC, but was much less so with GESAMP where the UN technical secretaries, the GESAMP Secretariat, and the Marine Environmental Protection Committee of the International Maritime Organization are mandated to move the information into the appropriate UN

decision making system(s). In general though, once a work is published by an intergovernmental body, attention moves rapidly to other assigned projects rather than allocating additional resources to advertise and disseminate its published work, and ensure that it is being used.

The limited attention to dissemination introduces several challenges. Application of best practices for distribution and promotion of new publications may be outside a group's general scope and interest. As a consequence, methods for tracing the use and influence of the group's information are rarely put to use. Making the new information more easily visible and interpretable and its significance more obvious for required decisions and policies is a second challenge. This matter is a translation issue lying at the intersection of environmental science, communication, policy, and management (Holmes and Clark, 2008; Tribbia and Moser, 2008). It is noteworthy that some intergovernmental organizations have produced communication products developed with their potential users clearly in mind, e.g., GOMC's *Gulf of Maine Times* and fact sheets; various Web sites of UN agencies, such as the United Nations Environment Programme; and reports of the IPCC. In some intergovernmental organizations, e.g., GOMC and IPCC, scientists often work closely with communication specialists and science translation writers. Some individuals or organizations work with policy makers directly and continuously, the best example being the IPCC teams of scientists and government policy writers. GESAMP did not follow this practice, except for Report and Studies No. 70, *A Sea of Troubles*, published in 2001. Clearly, if the growing marine environmental grey literature, now mostly published on the Web, is to become more noticeable, accessible, and useful, links between researchers and potential users need to be strengthened and the relevance of new information clearly expressed.

The use and influence of information in grey literature is also affected by the manner in which it is packaged and communicated. Many scientific technical reports are produced for specific purposes. GESAMP's reports, for example, are always solicited and funded by the agencies requiring certain information, as is also frequently the case with GOMC's. But often, the target audience of a grey literature publication is not identified clearly (Healy and Asher, 1995; Tribbia and Moser, 2008). This situation creates uncertainties in the flow of information: who is requesting the information, what is being requested, what is being produced, and how it is being used?

From our study of a FAO Regional Fisheries Body (Soomai, 2009), it became evident that tracing the flow of information among stakeholders in the management of a marine living resource can help in understanding the stages involved in the preparation, production, distribution, and use of specific, grey literature fisheries reports. Soomai (2009) demonstrated these stages in the case of the FAO/WECAFC Shrimp & Groundfish Working Group by using a structured questionnaire aimed at multiple stakeholders. In fisheries, scientific information on regional and international issues is generally produced by international organizations such as the FAO in collaboration with national scientists. A sizeable number of highly technical fisheries assessment reports is produced which continually

adds to the overall knowledge base. This information is useful within the fisheries scientific community. However, communicating results to administrators, managers, and the fishing industry is often a challenge due to the highly technical nature of the subject, and in this case study is a clear impediment to information use. Credibility of the grey literature appears to be dependent on the degree to which the entire range of stakeholders is included in the preparation of reports or is provided with the information output. Often, many stakeholders are not consulted (Soomai, 2009). Consequently, low recorded levels of usage can result in grey literature reports being deemed less credible and salient.

11.4 Conclusions

An underlying hypothesis of our research is that many of the problems currently facing the marine environment and its living resources could be solved or mitigated by better use of existing information, especially information published as grey literature by intergovernmental organizations, such as in our case studies, GESAMP, GOMC, and FAO. For some geographic areas, as Soomai's (2009) study of FAO in the Caribbean demonstrated, grey literature is the most comprehensive source of available fisheries scientific information. More generally, grey literature from intergovernmental bodies on marine environmental and fisheries questions is an increasingly significant component of the global knowledge base on these matters. But barriers to the use and potential influence of this literature persist, even with increasingly wide deployment of new sophisticated search engines. Finding what is needed at the appropriate time, whether it be a database, a primary paper or a technical report, remains a major problem in coastal and ocean affairs (M. Butler, personal communication). In other words, awareness remains a major barrier to the information's effective and widespread communication and use.

The use problem is clearly multifaceted. Factors influencing use include: a general misunderstanding of the credibility and value of information published as grey literature; the challenge of determining the relevance of particular information sources within an overwhelming volume of information; the wide range of publication options now opening up due to advanced digital technologies; and the scattered distribution or sources of those publications.

Questions about the credibility of information published as grey literature by intergovernmental bodies can act as a major barrier to its effective use. There is often the false assumption that grey literature is never or rarely impartially refereed (e.g., National Resources Canada 2007 definition of grey literature), in contrast to the refereed primary journal and monograph literature. Hence, members of the scientific community, managers, and policy makers may distrust the quality of this information, even though in many cases it has been rigorously reviewed prior to release, as in the case of the technical reports published by GESAMP, GOMC, and FAO. GESAMP's reports, as peer-reviewed literature, are particularly re-

spected in scientific and management circles, as are the technical reports of many scientific groups and agencies worldwide. For example, in Canada, many series of scientific reports published by the federal government departments of Environment Canada and Fisheries and Oceans are rigorously reviewed and edited. It is due time that this genre of publication gained the respect that it deserves, and that it not be considered substandard or secondary if issued by credible sources. However, in this scenario of credibility, the term "grey literature" may have pejorative connotations that create a barrier to its use. The term "technical report literature" is seen more positively even though it may be synonymous with "grey literature."

Doubts about credibility also arise when the technical language of scientific grey literature places it beyond the comprehension of audiences that could benefit from it. The complexity of the language feeds the credibility question and, as a result, use of the information may be impeded. Of course, the same can be said of "primary scientific literature," where the language in most fields of research has become increasingly specialized and opaque or incomprehensible to the non-specialist, including policy makers. This problem is readily resolved by employing communication specialists, as is done by the journal *Science* on a weekly basis, with easily understood summaries of key papers.

The relevance of grey literature addressing global environmental challenges, such as that produced by intergovernmental organizations in our studies, warrants research engaging the wider scientific and policy communities. The seriousness of global environmental conditions in the early 21st century demands interdisciplinary attention (Myers, 2009). Our ongoing research on "Environmental Information: Use and Influence" uses such an approach incorporating the fields of information management, marine science, environmental management, and fisheries resource management. This approach is leading to a greater understanding of information life cycles and barriers to the diffusion, use, and influence of scientific information. Insights about communicating the value of such information in grey literature to professional and public audiences are also evolving.

Our continuing research will consider other governmental and intergovernmental organizations in the marine environmental arena, test additional hypotheses about the life cycles of grey literature information, and check the validity of our principal findings and conclusions.

11.5 Recommendations

Our research to date of the publishing practices of three marine intergovernmental organizations suggests several ways to improve awareness, retrieval, use, and influence of their literature:

1. The target audiences to whom the publications are directed should always be considered and the publications written accordingly. Production of less technical publications, e.g., fact sheets and pamphlets, is a specific activity

that should be factored into the resource assessment and management process.

2. Visibility, awareness, reading, and use can be significantly enhanced by releasing the information in a variety of publication venues and formats.
3. Each publication should be described consistently to facilitate efficient information retrieval (e.g., the title of a publication and the authoring organization should be referred to in the same way).
4. Raising awareness can be accomplished through the inclusion of online publications and bibliographic citations in sources such as Web of Science, Web search engines, and subject specific databases. To improve retrieval, producers of grey literature and associated agencies can post online listings of their publications, and awareness of grey literature can be spread through publication announcements in newsletters and blogs.
5. Publicising new reports and provision of copies of reports to relevant indexing/abstracting agencies will increase the likelihood of greater awareness and use.
6. Web links can facilitate awareness of grey literature available in digital formats. The more online referrals a producer can obtain to its publications, the more likely they will be located and ultimately used. Similarly, awareness may increase by arranging for related organizations to host copies on their Web sites.
7. As information technologies advance, digital publication and dissemination should be utilized strategically to promote access and reading in environments reliant on on-line systems.
8. Effective use of grey literature can be achieved by users remaining current with advanced search engine technologies.
9. The capacity of research scientists to communicate scientific information to wider audiences can be enhanced through specialised training. As shown by GOMC, science translation specialists can be employed to prepare summary reports, fact sheets, and articles for newspapers directed at decision makers and the public. Communication strategies can address situations where physical access to information is still a challenge in spite of the availability of online sources and advances in Web-based technologies.
10. Information published online is often incorrectly regarded as unreliable, regardless of the source. Grey literature is especially vulnerable in this regard. Intergovernmental organizations, such as GESAMP, which employ rigorous internal and external peer-review and high editorial standards, could directly address credibility concerns by clearly noting the level of review to which their publications are subjected. Thus, use of this literature will be promoted by direct evidence of its authority and validity.

11.6 Acknowledgements

The research for this chapter was supported by the Social Sciences and Humanities Research Council of Canada, the Gulf of Maine Council on the Marine Environment, the International Maritime Organization (United Nations, London), Environment Canada – Atlantic Region, and Dalhousie University. We thank Michael Butler of the International Ocean Institute, Dalhousie University, for comments on this chapter during its preparation, and Julie Woods and Colleen Delaney, Dalhousie University graduate students, for their assistance in finalizing the text.

References

Bar-Ilan, J. (2008), What is h-index? A comparison of WoS, Scopus and Google Scholar, - In: Scientometrics, 74, pp. 257-271.

Bolin, B. (2007), A history of the science and politics of climate change. The role of the Intergovernmental Panel on Climate Change. Cambridge: Cambridge University Press. ISBN 978-0-521-88082-4.

Borgman, C.L. (2007), Scholarship in the Digital age: Information, Infrastructure, and the Internet. Cambridge, MA: The MIT Press.

Cordes, R. (2004), Is grey literature ever used? Using citation analysis to measure the impact of GESAMP, an international marine scientific advisory body. – In: The Canadian Journal of Information and Library Science, 28, pp. 45-65.

Cordes, R., B.H. MacDonald, P.G. Wells. (2006), Publications of the Gulf of Maine Council on the Maine Environment and Their Use. Halifax, Nova Scotia. Report to the Gulf of Maine Council on the Marine Environment.

http://www.gulfofmaine.org/council/publications/GOMCpublications-report-Cordes-2006.pdf

Cossarini, D.M. (2009), Unpublished GOMC interview data.

de Alwis, G, S. Majid, A.S. Chaudhry, (2006), Transformation in managers' information seeking behaviour: a review of the literature, - In: Journal of Information Science, 32 (4), pp. 362-377.

Evans, J.A., J. Reimer. (2009), Open access and global participation in science. In: Science 323, 20 Feb 2009, pp.1025.

GESAMP (IMO/FAO/UNESCO-IOC/WMO/WHO/IAEA/UN/UNEP Joint Group of Experts on the Scientific Aspects of Marine Environmental Protection). (2008), Publications. http://s244621454.onlinehome.fr/common.php?id=6

GOMC. (Gulf of Maine Council on the Marine Environment). (2007), Action Plan 2007-2012. www.gomc.org/publications

FAO (Food and Agriculture Organization). (2009), Regional Fishery Bodies Summary Description: Western Central Atlantic Fishery Commission. http://www.fao.org/fishery/rfb/wecafc/3/en

Healy, R.G., W Ascher. (1995), Knowledge in the policy process: incorporating new environmental information in natural resources policy making,- In: Policy Sciences, 28 pp. 1-19.

Holmes, J., R. Clark. (2008), Enhancing the use of science in environmental policy-making and regulation,- In: Environmental Science & Policy, 11, pp. 702-711.

Hutton, G.R.G. (2009), Developing an Inclusive Measure of Influence for Marine Environmental Grey Literature. Master of Library and Information Studies Thesis, Dalhousie University, Halifax, Canada.

IPCC (Intergovernmental Panel on Climate Change). (n.d.), History. http://www.ipcc.ch/organization/organization_history.htm

IPCC (Intergovernmental Panel on Climate Change). (2009), Progress report. Report of the Joint IPCC/WCRP/IGBP workshop: new science directions and activities relevant to the fifth assessment report of the Intergovernmental Panel on Climate Change, 3-6 March 2009, University of Hawaii, Honolulu, Hawaii.

http://www.ipcc.ch/meetings/session30/inf3.pdf

MacDonald, B.H, R.E. Cordes, P.G. Wells. (2004), Grey literature in the life of GESAMP, an international marine scientific advisory body, - In: Publishing Research Quarterly, 20 (1), pp. 25-41.

MacDonald, B.H, R.E. Cordes, P.G. Wells. (2007), Assessing the diffusion and impact of grey literature published by international intergovernmental scientific groups: the case of the Gulf of Maine Council on the Marine Environment, - In: Publishing Research Quarterly, 23 (1), pp. 30-46.

McNie, E.C. (2007), Reconciling the supply of scientific information with user demands: an analysis of the problem and review of the literature,- In: Environmental Science & Policy, 10, pp. 17-38.

Myers, S.S. (2009), Global Environmental Change: The Threat to Human Health. Worldwatch Report No. 181. Washington, DC: Worldwatch Institute. ISBN 978-1-878071-92-7.

Natural Resources Canada. (2007), From Impacts to Adaptation: Canada in a Changing Climate 2007. Glossary. http://adaptation.nrcan.gc.ca/assess/2007/index_e.php

Palfrey, J., Gasser, U. (2008), Born Digital. Understanding the First Generation of Digital Natives. New York: Basic Books.

Pravdić, V. (1981), GESAMP, the First Dozen Years. Geneva: United Nations Environment Programme.

Soomai, S.S. (2009), Information and Influence in Fisheries Management: A Preliminary Study the Shrimp and Groundfish Resources in the Brazil-Guianas Continental Shelf. Master of Marine Management Research Project, Dalhousie University, Halifax, Canada.

Thatje, S., J. Laudien, O. Heilmayer, C.E. Nauen. (2007), Understanding El Nino – The importance of grey literature in coastal ecosystem research and management. – In: Marine Policy, 31 (2), pp. 85-93.

Tribbia, J., S.C. Moser. (2008), More than information what coastal managers need to plan for climate change, - In: Environmental Science & Policy, 11, pp. 315-328.

Vaughan, L., D. Shaw. (2005), Web citation data for impact assessment: a comparison of four science disciplines, - In: Journal of the American Society for Information Science and Technology, 56, pp. 1075-1087.

Voss, B., J. Webster. (2007), North Pacific Marine Science Organization (PICES) Publications: What's the Access and Where's the Value?

http://darchive.mblwhoilibrary.org

Wells, P.G. (2003), State of the marine environment (SOME) reports – a need to evaluate their role in marine environmental protection and conservation, - In: Marine Pollution Bulletin, 46, pp. 1219-1223.

Wells, P.G., R.A. Duce, M.E. Huber. (2002), Caring for the sea — accomplishments, activities and future of the United Nations GESAMP (the Joint Group of Experts on the Scientific Aspects of Marine Environmental Protection), - In: Ocean and Coastal Management, 45, pp. 77-89.

Wells, P.G., T. Höfer, M. Nauke. (1999), Evaluating the hazards of harmful substances carried by ships: The role of GESAMP and its EHS working group, - In: Science of the Total Environment, 237/238, pp. 329-350.

Wells, P.G., B.H. MacDonald, R.E. Cordes, D.M. Cossarini, G.R.G. Hutton, J.L. Woods. (2009), Use and influence of marine environmental information: continued studies of the publications of intergovernmental groups such as the Gulf of Maine Council on the Marine Environment. Poster presentation, Gulf of Maine Symposium, Advancing Ecosystem Research for the Future of the Gulf, St. Andrews, New Brunswick, 4th-9th October, 2009.

Windom, H.L. (1991), GESAMP: Two decades of accomplishments. London: International Maritime Organization.

Chapter 12

Grey Literature in Karst Research: The Evolution of the Karst Information Portal (KIP)

Todd A. Chavez, University of South Florida, USA

12.1 Introduction

The Karst Information Portal (KIP) is a digital library linking scientists, resource managers, and explorers with quality information resources concerning karst. Beginning in 2006 as a partnership between the University of South Florida Libraries, the National Cave & Karst Research Institute, the University of New Mexico Library, and the Union Internationale de Spéléologie (UIS), the KIP initiative has expanded to include databases concerning cave minerals, speleothem dating, and coastal cave surveys. This chapter outlines the project's evolution and describes efforts to improve information access and preservation for karst researchers, a globally distributed research community characterized by a highly interdisciplinary knowledge base often drawn from and memorialized in grey literature.

12.1.1 What is Karst?

Karst is a globally distributed terrain resulting from the dissolution of soluble rocks, such as limestone and dolomite. This dissolution occurs when rain water infused with carbon dioxide passes through layers of soil and bedrock (see Figure 1). Karst regions contain aquifers and geological structures, such as sinkholes, springs, and caves, many rare and endangered species, as well as significant archaeological and paleontological resources (Culver et al. 2000; Culver et al. 2001; LaMoreaux 2005; Northup, et al. 2003; Straus 1979).

Globally, approximately 1.6 billion people depend upon the health of karst terrains and aquifers for drinking water (Drew and Hotzl 1999; Ford and Williams 2007). Geologic hazards in karst cost billions of dollars each year (Cobb and Currens 2001), yet karst is the least studied and most vulnerable type of terrestrial landscape (Williams 1993). The full potential of karst for benefit or hazard to the global ecosystem, including humanity, remains poorly understood. The karst

research community seeks to facilitate better, science-based management practices in karst terrains worldwide (Veni et al. 2001).

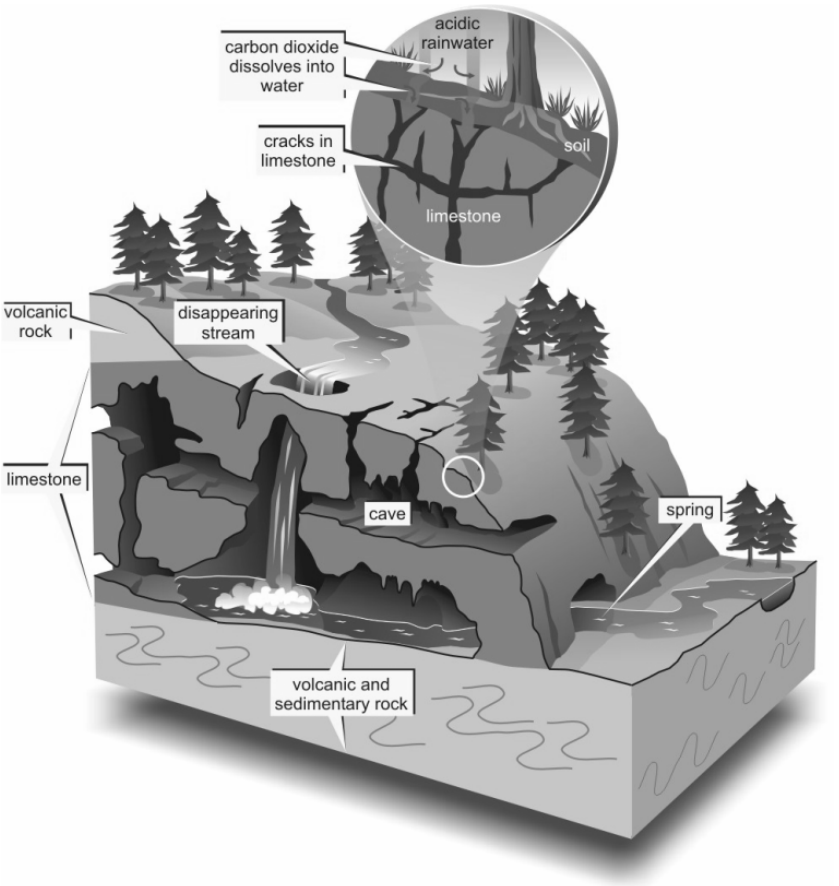

Figure 1. Karst Terrain (Natural Resources Canada)

12.2 Project Background

In 2005, an interdisciplinary group of faculty, librarians, and graduate students at the University of South Florida (USF) met to discuss global information needs. This group prioritized meeting both institutional and community challenges facing water resource managers and more specifically those concerned with karst terrains, a complex and vulnerable type of geologic landform common throughout Florida. The group decided to construct an information portal centering on karst

research that the USF Libraries would host and maintain in collaboration with related academic departments.

In January 2006, a group of 29 scientists, information specialists, and policy makers representing 18 organizations from across the globe met in Carlsbad, New Mexico, to explore development of the Karst Information Portal (KIP) to serve as a repository for karst information, to advance collaboration among the international community of karst researchers, and to promote knowledge discovery through innovative applications of metadata.

12.2.1 Who is Engaged in Karst Research?

Karst researchers come from a variety of established disciplines, including anthropology, biology, chemistry, geology, and geography. Engineers have a pressing need to understand karst environments in order to mitigate geohazards, such as sinkholes, and to appropriately manage water resources. Space scientists' recent interest in the potential for extraterrestrial caves to shelter and support Mars explorers led to increasing focus on terrestrial cave environments. Finally, land use and resource managers as well as policy makers depend on the work of this diverse cadre of scientists for an appropriate foundation for best management practices in karst terrains.

All of these scientists bring unique discipline-based theoretical frameworks and methodologies to the challenge of understanding karst environments. USF Department of Geography Chair Robert Brinkmann summarized the challenge to the consumer of this important research during his 2007 presentation to the International Congress on Karst Hydrogeology & Ecosystems:

"The karst community and its knowledge base are in some ways similar to French cheese – and not because of the amount of time each spends in caves. Rather it is a fragmented community, often identified with a region or a specific cave or discipline. And so goes the information that services the research. Given the example of the US wherein we typically limit ourselves to consuming just four varieties of French cheese, how can we learn about the other types without affirmative efforts to share and collaborate in research and in the production, dissemination, and preservation of karst information resources?"

12.2.2 What Does Karst Research "Look" Like?

Historically, early karst research fell into a "marginal" category of scientific inquiry and has only recently migrated from those margins of traditional disciplines to positions of more central concern and inquiry (Veni 2007). Today, the karst research community's information environment is as highly inter-disciplinary as those engaged in the research. An understanding of this

environment is an important part of any attempt at a comprehensive solution to an array of global social, environmental, and health challenges.

NCKRI Executive Director George Veni describes the pre-1950 period as the "Curiosity" phase during which the term "karst" was unknown; minimal data concerning the geomorphology and hydrology of caves and karst existed; caves were curiosities found on the remote margins of human population centers; and karst aquifer behavior was at variance to known hydrologic principles. Beginning in 1950 and extending to roughly 1980, the "Experimentation" period saw significant increases in the quantity and diversity of cave- and karst-data collected through exploration and surveying activities, although scientific understanding of the emerging concepts were largely confined to a small cadre of individuals. One of the reasons for the marginalization of karst research concerns the involvement of non-academics in data collection and exploration. Like ornithology and astronomy, cave and karst research benefits from involvement of a significant number of "amateurs" who are passionate about caving and preservation of these resources (Palmer 1996). Many of these individuals joined the ranks of the academically trained scientists in the 1950s, thereby increasing interest in karst research topics.

A significant proportion of the information produced during the "Curiosity" and "Experimentation" periods found its way into the grey information realm. Few established scientific journals published these dedicated amateurs' work, so their findings were reported in such grey channels as cave club ("grotto") newsletters, personal reports circulated to a narrow audience, and vertical files in organization offices.

Veni refers to the years 1980-2009 as the "Application" period, a time when karst became broadly known – perhaps as a result of pressures to improve land and water resource management and to understand climate change – but remained poorly understood. Karst topics more frequently began to appear at non-karst conferences and increasing numbers of non-caving scientists became interested in karst environments and published their work in the "white" literature. This was confirmed in a recent study of four widely-used scientific indices, where searches using 632 karst-related terms determined that, over the period 1980-2005, publication on cave and karst themes increased substantively (Florea, Fratesi, and Chavez 2007).

Because karst researchers are faced with discovering and evaluating relevant information sources and obtaining and preserving "grey" karst information sources, an online, open-access portal that contained grey information useful to karst researchers was suggested and a needs assessment performed.

12.3 Defining the Karst Information Portal's Grey Information Mission

Modern geological research depends as much, if not more, on previously known information than on new data. Yet, a significant proportion of the findings generated by formal and informal research fail to find its way into the published scientific literature, instead becoming part of the growing body of grey information. Grey information is a critical research element in many scientific disciplines (see Aina 2000; Bichteler 1991; Cordes 2004; Dunn 2004; Gelfand 1998; Hanner 1990; Musser 2003; Noga 2004; Sulouff et al. 2005; Trivelpiece et al. 2000; Weintraub 2000), and thus its importance for karst researchers must be accurately assessed.

12.3.1 Planning the Information Needs Assessment

In 2006, information specialists from the University of South Florida Libraries and the School of Library and Information Science planned and conducted a global information needs assessment for the KIP. The survey elicited responses in three information need categories: 1) information content (e.g. format, subjects, and organization); 2) services (e.g. blogs, newsfeeds, and tagging services); and 3) research tools (e.g. data-mining and computational utilities) (Chavez, et al. 2007).

Respondents received the definition of grey literature adopted during the Third International Conference on Grey Literature:

"[T]hat which is produced by government, academic, business, and industries, both in print and electronic formats, but which is not controlled by commercial publishing interests and where publishing is not the primary activity of the organization." (Farace 1998)

The phrase "non-refereed and self-published documents generated by speleological groups and other non-governmental groups/individuals such as expedition reports," was appended to the core definition to accommodate known grey information types of specific relevance to the karst community.

12.3.2 Consuming, Producing, and Accessing Grey Information Concerning Karst

Respondents (n=66) reported heavy consumption and production of grey information resources, with 96.6 percent reporting regular use of a variety of the 46 grey information formats listed in the survey. Conference proceedings/papers, trip and cave reports, theses/dissertations, and maps in any format were the most common. Responses to subsequent questions identified the aforementioned as the most commonly produced grey information resource, plus images, records of speeches or invited talks, and research proposals. All respondents reported

difficulty locating all grey information types (except audio files), and the survey results reflected a strong correlation between the grey information types consumed/produced and difficulty locating same. For example, even though 86.2 and 80.8 percent of respondents reported consuming and producing conference papers, respectively, 47 percent reported difficulty in locating this information (Chavez et al. 2007, 9-11).

In terms of grey information consumption and production relevant to karst research, academic researchers account for a significant percentage (74.1) of respondents who report producing grey information in some format. Roughly 69 percent of the researchers contribute to conference proceedings, deliver speeches/invited talks, or generate images, while 55 percent produce trip and cave reports, and 51.7 percent create or contribute to cave registries or entrance databases. Reflecting the important role that non-academics play in karst exploration and research, the study concluded that 84.6 percent of self-identified cavers report producing grey literature, with trip and cave reports and cave registries or entrance databases the most frequent contributions. Responses also indicate that five of the six college or university student respondents produce grey information, including conference papers, theses/dissertations, trip and cave reports, images, datasets, and maps (Chavez et al. 2007, 8).

The survey confirmed previous usage studies as well as anecdotal reports, and it demonstrated that trip and cave reports are a significant special case of grey literature for karst researchers and cavers. This finding illustrates the importance of studies that focus on specific knowledge domains (e.g. Bichtler 1991, Corbett 1989, Derksen and Sweetkind-Singer 2003, Haner 1990, and Walcott 1990). Commonly called geological field trip books, trip and cave reports are produced by local experts to support excursions into specific field locations. The reports typically include coverage of transportation resources and relate information about local cultural, geological and geographic features, and conditions at a specific point in time (Bichtler 1991, 41-42). Both grey and "white" publications often contain citations to trip and cave reports or field books, but, because they are often published by organizations lacking an infrastructure to facilitate wide distribution, librarians are hard pressed to acquire copies, and, once in hand, cataloging is a challenge (Haner 1990, 166-7; Walcott 1990).

12.3.3 Archiving Grey Information for Karst Researchers

Eighty-nine percent of the respondents to the survey reported that they produce grey information in some form, but 28.3 percent do not formally archive their output (Chavez et al. 2007, 11-12). Despite the clear need for a systematic archival and preservation strategy, the survey revealed an important consideration as the KIP emerged: data sensitivity. Respondents to the survey and participants in all of the subsequent presentations concerning the KIP initiative have expressed concern for the security of cave entrance location data and water-tracing information.

12.3.4 Additional Findings

Several survey questions focused on KIP's potential for promoting collaboration via services or capabilities, such as file sharing, RSS feeds, blogs, data management tools, web indexing, and directory services. It is notable that, even with the advent of Google, a large percentage of respondents continue to rank searchable link collections and search tools (83.9 and 60.7 percents respectively) as important for inclusion in the portal. This suggests that domain-specific information portals are valuable tools for information discovery in specialized areas such as karst research.

Respondents generally supported social networking services to improve connections among those interested in karst issues. Non-academics generally supported but academic resisted allowing KIP managers to serve as evaluators of grey information . Regardless of their feelings about social networking or evaluation services, respondents indicated that grey information should be a key function of the portal, with 99 percent considering grey literature's inclusion very or somewhat important and 96 percent rating grey literature digitization very or somewhat important (Chavez, et al, p. 14). These conclusions informed the KIP's design as well as collection building priorities.

12.4 The Karst Information Portal

The Karst Information Portal (KIP) went online in June 2007 at www.karstportal.org. As of October 2009, KIP's electronic collection contains 4,756 records. These items include 28 distinct document types, including images, maps, grey literature works, peer-reviewed journal articles, and raw data organized into databases. To remain current with developments in the field, graduate students in the geosciences and library and information science assigned to the KIP initiative systematically scour the Internet for karst-related resources. Researchers and authors are encouraged to upload their own work. This happens less frequently, but noteworthy content submissions deposited in this manner have occurred.

Access to KIP is freely available to the public; however, registration is required to take advantage of certain features and benefits. As of October 2009, 137 individuals from 12 different countries have registered. Upon registration, users are asked to indicate their areas of research interest or specialization. To date, users have self-identified 30 distinct areas of karst focus.

In 2007, KIP project partners met with representatives of the Union Internationale de Spéléologie (UIS), the international body for caving and speleology, to formalize an international partnership. As a result of the relationship, UIS members now collaborate on website design and governance as well as contribute to the growing collection of information resources.

Project planners initiated KIP's "soft-launch" in June 2007, with 1) a collection of nearly 3,000 bibliographic references to key karst information sources; 2) a

small collection of scanned electron photomicrographs; 3) social networking applications, including RSS feeds, news services, community forums for online discussions, and a directory of relevant organizations and registered users; and 4) a collection of images and oral histories of key karst researchers. In July and August 2007, the project partners attended one national and three international conferences, in all cases presenting papers and leading discussions concerning KIP and its mission. User registrations increased 74 percent following these meetings.

In October 2007, a 1.0 FTE faculty line, designated "Assistant in Karst Information Management," joined the KIP team to drive portal expansion. The incumbent was a recent graduate of the doctoral program in geography with specialization in karst hazards and land use. In January 2008, a graduate student in the field augmented the faculty position. Both lines fell victim to a budget reduction in late 2009, but the project continues with support from a librarian.

12.4.1 The KIP Content Collection

The KIP collection emphasizes grey information and retrospectively digitized content from both the grey and white information realms. This strategy provides a valuable service to the karst community, given the considerable effort expended in pursuit of primary sources. It also alleviates the issue of effort duplication that occurs when researchers tackle problems that have, unbeknownst to them, already been addressed by other. This lack of awareness is often the direct result of important karst literature's inaccessibility.

Table 1 describes the collection in detail. The left column lists currently held information formats. The "Records" column details the number of records in that format. The "Digital Objects" column refers to the number of those records that link to internally held digital objects. The "Grey Information" column records the number of the records that meet the established definition of grey information (see Farace above) regardless of whether the item is locally held or available in print.

The preponderance of monograph records and the limited selection of locally held digital objects reflect KIP's initial upload strategy. A bibliographic database created by karst scholar and KIP partner Diana E. Northup for her monograph *A Guide to the Speleological Literature of the English Language, 1794-1996* was subsequently donated to the project and helped "jump-start" the collection. Although locally held digital content is the preferred strategy, the survey encouraged KIP's contributions to information discovery and bibliographic control. On a national and global level, important information resources essential to karst research are elusive. Sometimes their existence is unknown outside a small circle of karst researchers. The most comprehensive collections are usually in private hands and are generally focused on one or two specialized karst topics, regions, or features of primary interest to the collector.

Table 1. Characteristics of the KIP Content Collection, October 2009.

Information Formats	Records	Digital Objects	Grey Information
Monographs	2774	14	817
Articles	624	58	0
Serials (Analytic)	364	312	109
Technical Reports	370	112	312
Newsletters	148	148	148
Trip & Cave Reports	126	57	126
Archival Materials	74	0	74
Proceedings	74	35	52
Internet Resources	51	0	0
Theses & Dissertations	40	3	40
Maps & Cartographic Materials	38	11	26
Databases	22	3	22
Oral Histories	14	14	14
Article Preprints	8	8	0
Visual Materials / Images	7	7	7
Book Chapters	7	0	0
Bibliographies	5	4	4
Power Point Presentations	4	4	4
GIS Data	3	0	3
Microforms	1	0	1
Speeches & Invited Talks	1	0	1
Computer Software	1	0	1
TOTALS	**4,756**	**790**	**1,761**

12.4.2 Areas of Collection Emphases

In 2009, citation analysis, institutional research intensity (USF), and estimates of potential community impact identified three areas of collection emphasis both within KIP and to support a more comprehensive library collection initiative.

Karst Hydrology. Karst aquifers provide drinking water to between one and two billion people worldwide (Veni et al). Groundwater contained in these aquifers is easy contaminated because surface water receives no filtering from the hard limestone bedrock as it rapidly makes its way downward to recharge the

aquifer. For this reason, any contaminants or pollutants on the surface are rapidly washed into the groundwater. This can have serious public health implications for populations relying on that groundwater for drinking water supplies, especially in developing areas that lack strong health care and water utilities infrastructures. Greater understanding of these complex systems can lead to more effective contamination mitigation strategies and technologies for karst aquifers.

Paleoclimatology. Carbonate rocks often contain important clues to localized climate conditions in the distant past. Many caves are isolated or difficult for the average person to access and therefore can be particularly valuable sources of unspoiled paleoclimate data. Samples extracted from stalactites and stalagmites are less subject to influence by outside forces than data collected from, for example, sediment cores taken from lake beds. Understanding how and why climate conditions changed in the past can help identify the best means to address current and future climate change issues.

Policy Innovation and Development. Policy solutions for karst-related environmental and public health issues have been implemented in various locations throughout the world; however, this is an underdeveloped sub-field of karst studies. There are many locations where such solutions likely would be appropriate but have never been attempted. Even for those locations where policy-driven approaches are taken, those approaches vary wildly in structure and regulatory strength. There is currently no universally accepted approach to policy development with regards to karst and the human activities that threaten it. By making policy innovation a collection priority, the USF Libraries help facilitate education on the importance and feasibility of policy-based approaches generally, as well as the development of specific policy-based techniques for managing karst lands.

12.4.3 New Collection Directions

Since the 2007 "soft-launch," collection building has expanded in five directions, with early emphasis on serials, bibliographies, oral histories, database development, software applications, and modeling and research.

Serials. The National Speleological Society has emerged as a strong KIP partner and source of important information, both grey and white. The USF Libraries digital collections unit has digitized entire runs of the *NSS News* (1958 to present), the *Bulletin to the National Speleological Society* (1940-1958), and select issues of the *SpeleoDigest*. Future plans call for completing the *SpeleoDigest* back files, exploring ebook publication, and incorporating the "NSS Volunteer Value Database" in KIP.

KIP digital serial content includes the Association for Mexican Cave Studies newsletters (full runs of three distinct titles), *Espeleo Informe Costa Rica*, *GEO2* (in progress), *Helictite: Journal of Australasian Speleological Research* (in progress), the *Proceedings of the International Symposium on Vulcanospeleolgy*,

and the *Proceedings of the National Cave and Karst Management Symposia*. As of October 2009, negotiations to host the *International Journal of Speleology* and the *Journal of Cave and Karst Studies* as open-access journals are in the final stages. The journal *Studia Universitatis Babes-Bolyai Geologia* is already part of the collection. The later established scientific journals have become distinctive components of the KIP collection because of their value to karst researchers and the potential benefits created by joining forces to limit costs and raise visibility.

Bibliographies. In 2008, KIP project managers initiated discussions with the creators of three important bibliographic resources concerning karst: the *Bulletin Bibliographique Spéléologique = Speleological Abstracts* (Centre de Documentation UIS), the Texas Speleological Survey Bibliographic Database, and *Speleogenesis*' KarstBase database. A merger should permit each entity to meet stated goals, with KIP providing an organizational and infrastructure "umbrella" to facilitate those activities. Participants in these important endeavors report pressures that, unless alleviated, could endanger their survival –print publications' sky-rocketing costs, technology migration pressures, and long-term preservation. KIP was established to manage these pressures within the workflows of an academic library collection. KIP presents a viable alternative to publication cessation or to taking these projects commercial.

Karst Oral Histories. In conjunction with the USF Libraries' Oral History Program, KIP managers conducted oral history interviews with leading names in a variety of karst science fields, including exploration, cave mapping, and applied ecology. The karst oral history project seeks to preserve for future researchers the experiences, thoughts, and insights of prolific karst researchers and authors Alexander Klimchouk, Derek Ford, and William and Elizabeth White. The complete audio recordings of these interviews are available for download via KIP, along with a written transcript for each.

Database Development. The karst researchers require increased capacity to create databases relevant to their areas of study. The infrastructure to support these resources must be user-friendly, established on best practices/standards, powerful, capable of efficient/unmediated data exchange, and archived for future access. On numerous occasions since KIP's launch, potential partners have approached the project team to solicit input and assistance in designing and implementing novel databases relevant to karst research. A selection of specific examples illustrates the need for this capacity.

The Cave Mineral Database (CAMIDA) is a collaborative project of the USF Libraries, UIS's Cave Minerals Commission, the Karst Information Portal, "Emil Racoviţă Institute of Speleology (Romania), and the Karst Research Group at University of South Florida. CAMIDA is an open-access collection of geological, mineralogical, crystallographical, and protection/conservation information on all minerals discovered in caves around the world.

Professor Donald McFarlane (Scripps College) is collaborating with KIP Project Manager Todd Chavez and others to create *The Bibliography of Speleothem Research*, an archive of peer-reviewed speleothem research papers

specifically intended to be searchable by geographic and/or geochronological parameters.

Professor John Mylroie (Mississippi State University, Department of Geosciences) and caver Mike Lace propose collaboration to create a database and repository of all known information on sea caves and dissolution caves in coastal settings and to make these data web accessible.

Future projects include a dye-trace database for the eastern United States (Karst Waters Institute); a joint collaboration to migrate the National Karst Map to a web-accessible database (USGS); a digital world karst map (USGS and the World Wildlife Fund); a karst geo-wiki to serve as the basis for informal science education and community participation (USF Professor Robert Brinkmann and collaborators from the National Park Service and the University of New Mexico); a joint archive of SEM images that supports user commenting (University of New Mexico); and a database of isotope data for southeast European cave fossils (USF).

Software (Freeware) Applications. As in many "small-science" research areas, individual karst researchers are often required to develop "just-in-time" software applications to support their work, usually without specialized training or concerns for future usability/functionality. In concert with database development described above, KIP project managers plan to develop web-accessible freeware software applications to facilitate karst-related research.

Karst Modeling and Research. Web-accessible scientific modeling tools (statistical, geospatial, etc.) that can efficiently incorporate and manipulate data resident in the USF Libraries' karst databases are natural extensions of the current collections. Using USF Libraries-developed software applications, users can collect and organize data subsequently imported into a USF Libraries-developed database and extracted to be included in models developed using USF Libraries-developed modeling tools.

Similar efforts to develop these capacities in the geosciences include two NSF-funded projects GEON (volcanology, seismology) and CHRONOS (earth history) – the leaders of both projects were early contributors to the KIP planning process. The projects represent significant advances in creating an integrated cyberinfrastructure serving the geosciences, and their experiences help guide KIP collection directions.

12.4.4 Services and Programming

Collections cannot exist in a vacuum. Context is important and contributes significantly to collection visibility and use. To that end, the initiative's strategic plan includes developing public programs, facilitating scholarly communication around KIP, and developing instructional collaborations.

In the long-term, the health of karst environments is dependent on enhancing understanding of karst environments among researchers outside of the informal

karst community and among members of the public -- from K-12 teachers to politicians and homeowners. Although the initiative is strongly digital in emphasis, non-web public programming facilitates the overall goal of increasing the impact of karst research and KIP's visibility. Partnerships with museums, television stations, and K-12 educators can serve to promote public understanding of karst environments. These avenues are being pursued, with the first radio and television spot highlighting the KIP and affiliated faculty due to air in November 2009. A YouTube video called "What is Karst?" was produced and posted in 2008, and as of October 2009 the video was viewed over 1,500 times.

Hosting conferences relevant to karst research also should increase KIP's visibility and impact. The first such conference is scheduled for May 2010. Members of the KIP team are formalizing instructional collaborations and course offerings that combine librarians and faculty from relevant academic departments to give graduate students hands-on experience with the concepts, techniques, and tools of library and information science relevant to their particular thesis and dissertation research. A recent collaboration involving the USF Libraries and Department of Geology can be replicated and expanded to include additional disciplines.

12.5 Conclusion

The Karst Information Portal grew out of a sense of the importance of grey literature to karst researchers and consumers of that research. Consistent with Professor Irwin Weintraub's oft quoted article, "The Role of Grey Literature in the Sciences,"

"In a world in which free trade and instantaneous communication have eliminated many of the barriers to information flow, grey literature is gaining greater importance as a source of information for much of the world's population. It is an indispensable resource for an informed and enlightened public and will undoubtedly continue to serve as a necessary supplement to journal literature well into the future."

An information needs assessment conducted by USF researchers confirmed this assertion and the use of the KIP since implementation supports Weintraub's general characterization in the specific case of the interdisciplinary domain of karst science. Geoscientists generally, and karst research specifically, regularly use and (less frequently) cite grey literature (Butkovich and Musser 1994). Interdisciplinary research domains, including library and information science (Aina 2000), the health sciences (Alberani et al. 1990; Dunn 2004), marine and fisheries science (Cordes 2004), economics (Mili 2000), and transportation studies (Osif 2000), increasingly reflect intense use of grey literature, though not to the exclusion of traditional published research. The pattern is clear.

Other conclusions drawn from the study were not well supported in the intervening two years. Analysis of KIP usage patterns since implementation has

necessitated reconsideration of the resource's social networking applications and community aspects. The discussion forums, in particular, have generated little interest, and user registration lags behind use. According to a report tracking usage during one six-month period, KIP was serially accessed by 189 different users from eight countries, but the directory only includes 137 registered users. All of these uses, with 14 exceptions, were tracked as coming from Google searches, a positive development that demonstrates success in efforts to increase the visibility of karst research content via KIP.

In the fall 2009, USF Libraries' personnel began to migrate the existing Karst Information Portal content to a new infrastructure. The previous architecture was visually appealing, and the content management system supported most basic metadata requirements, but refinements were needed. A decision to adopt the NSF-funded National Science Digital Library's Collection Workflow Integration System (CWIS) as the KIP's new content management system followed extensive testing and comparisons of several alternatives. A further decision to cease the resource's forums and other community aspects and to focus on KIP's digital library characteristics followed.

In the information needs assessment report, the authors suggested that, "When implemented, the KIP can serve as a model for similar studies of global interdisciplinary communities and the gathering and synthesis of literature to support the research needs of that community" (Chavez et al. 2007, 16). Events during the two years since KIP's implementation have emphasized the value of a library-led collaboration with global research communities.

Acknowledgements

I appear as the sole author of this chapter, but the project it describes represents a collaborative effort, with valued contributions by University of South Florida (USF) professors Robert Brinkmann, H. Len Vacher, Anna Perrault, Bogdan Onac, Philip Reeder, and Philip Van Beynan. Former KIP Manager E. Spencer Fleury and USF librarians Matt Torrence, Cheryl McCoy, Pete Reehling, and Ardis Hanson provided input into and assistance with a business plan describing the larger USF Libraries Karst Science Collection Initiative. Spencer has been a valued partner to me throughout the project. The USF Dr. Kiran C. Patel Center for Global Solutions' Mark Amen and Rebecca Harris were early catalysts in the project's "visioning" phase and supporters throughout. USF doctoral student Beth Fratsi and former USF students and current Western Kentucky University professors Lee Florea and Jason Polk worked on the project and continue to promote its success from afar. Both former and current NCKRI Executive Directors Louise Hose and George Veni, respectively, were there at the beginning and have each worked tirelessly to make the project real to the karst research community. Drs. Diana Northup (University of New Mexico) and Penelope Boston (New Mexico Tech) provided early vision and sustained support for the

project and for me personally. Among our many international colleagues, Peter Matthews and Alexander Klimchouk of the Union Internationale de Spéléologie stand out as perennial KIP supporters and visionaries. Finally, my colleague Mark Greenberg gave this chapter a thorough "once-over" as only he can and it has benefited from his expertise. All are valued colleagues, consummate scientists, and library supporters. None are responsible for errors that may have crept into this work.

References

Aina, L. O. "Grey Literature and Library and Information Studies: A Global Perspective." *International Journal on Grey Literature* 1.3 (2000): 179-82.

Alberani, Vilma, Paola De Castro Pietrangeli, and Anna Maria Rossi Mazza. "The Use of Grey Literature in Health Sciences: A Preliminary Survey." *Bulletin of the Medical Library Association* 78.4 (1990): 358-63.

Bichteler, Julie. "Geologists and Gray Literature: Access, Use, and Problems." *Science & Technology Libraries* 11.3 (1991): 39-50.

Butkovich, Nancy J., and Linda R. Musser. "Analysis of References Cited in US Geological Survey Bulletins, Circulars, and Professional Papers," in *Finding and Communicating Geoscience Information: Proceedings of the 28th Meeting of the Geoscience Information Society, October 25-28, 1993, Boston, Massachusetts*, ed. Connie Wick, 167-171. Alexandria, VA: Geoscience Information Society, 1994.

Chavez, Todd A., Anna H. Perrault, Pete Reehling, and Courtney Crummett. "The Impact of Grey Literature in Advancing Global Karst Research: An Information Needs Assessment for a Globally Distributed Interdisciplinary Community." *Publishing Research Quarterly* 23.1 (2007): 3-18.

Cobb, J., and J. Currens. "Karst: The Stealthy Hazard." *Geotimes: News Magazine of the Earth Sciences* (May 2001), http://www.geotimes.org/may01/feature2.html.

Corbett, Robert G. "Field Trip Guidebooks Need Not Be Gray Literature," in *Individual Workstations: Information Supermarkets for Geoscientists*, ed. Joan V. Lerud, 113-22. Alexandria, VA: Geoscience Information Society, 1989.

Cordes, Ruth. "Is Grey Literature Ever Used? Using Citation Analysis to Measure the Impact of GESAMP, an International Marine Scientific Advisory Body." *The Canadian Journal of Information and Library Science* 28.1 (2004): 49-67.

Culver, D.C., L. Deharveng, J. Gibert, and I. D. Sasowsky, eds. *Mapping Subterranean Biodiversity*. Karst Waters Institute Special Publication 6. Charles Town, WV: Karst Waters Institute, 2001.

Culver, D.C., L. L. Master, M. C. Christman, and H. H. Hobbs, III. "Obligate Cave Fauna of the 48 Contiguous United States." *Conservation Biology* 14.2 (2000): 386-401.

Derksen, Charlotte R. M, and Julie Sweetkind-Singer. "Accessing and Preserving Field Maps and Notebooks: The Stanford Geological Survey Map and Field Notebook Project," in *New Heights in Geoscience Information: Access and Technology: Proceedings of the 37th Meeting of the Geoscience Information Society, October 27-31, 2002, Denver, Colorado, USA*, ed. Lisa G. Dunn, 17-23. Alexandria, VA: Geoscience Information Society, 2003.

Drew, D., and H. Hotzl, eds. *Karst Hydrogeology and Human Activities: Impact, Consequences, and Implications*. Rotterdam: International Association of Hydrogeologists, 1999.

Dunn, Kathel. "Impact of the Inclusion of Grey Literature on the Scholarly Communication Patterns of an Interdisciplinary Specialty," in *Work on Grey in Progress: Sixth International Conference on Grey Literature, New York Academy of Medicine, NYAM Conference Center, December 6-7, 2004, New York, N.Y., USA*, 50-54. Amsterdam: TextRelease, 2004.

Farace, D.J. "Forward." *Perspectives on the Design and Transfer of Scientific and Technical Information: Third International Conference on Grey Literature, Jean Monnet Building, November 13-14, 1997, Luxembourg*, ed. D.J. Farace. Amsterdam: TransAtlantic GreyNet, 1998.

Florea, Lee, Beth Fratesi, and Todd Chavez. "The Reflection of Karst in the Online Mirror: A Survey within Scientific Databases, 1960-2005." *Journal of Cave & Karst Studies* 69.1 (2007): 229-236.

Ford, D.C., and P. W. Williams. *Karst Geomorphology and Hydrology*. Rev. ed. John Wiley & Sons, Ltd., 2007.

Gelfand, J. "Teaching and Exposing Grey Literature: What the Information Profession Needs to Know: Examples From the Sciences," in *Third International Conference on Grey Literature: GL '97 Conference Proceedings: Perspectives on the Design and Transfer of Scientific and Technical Information, November, 13-14, 1997*, 242-250. Amsterdam: TransAtlantic, 1998.

Haner, Barbara E. "Guidebook Citation Patterns in the Geologic Journal Literature: A Comparison between 1985 and 1967," in *Frontiers in Geoscience Information: Proceedings of the Twenty-Fourth Meeting of the Geoscience Information Society, November 6-9, 1989, St. Louis, Missouri, USA,* ed. Mary B. Ansari, 159-69. Alexandria, VA: Geoscience Information Society, 1990.

LaMoreaux, Philip. "Foreward," in *Encyclopedia of Caves*, eds. David C. Culver and William B. White, xvii-xviii. Amsterdam: Elsevier Academic Press, 2005.

Mili, Fethy. "Trends in Publishing Academic Grey Literature: Examples from Economics." *International Journal on Grey Literature* 1.4 (2000): 157-66.

Musser, Linda R. "What Geoscientists Cite: A Comparative Study," in *New Heights in Geoscience Information: Access and Technology: Proceedings of the 37th Meeting of the Geoscience Information Society, October 27-31, 2002, Denver, Colorado, USA*, ed. Lisa G. Dunn, 5-10. Alexandria, VA: Geoscience Information Society, 2003.

Natural Resources Canada. "What is Karst?" Geoscape Nanaimo: Geoscience for Central Vancouver Island Communities.

http://geoscape.nrcan.gc.ca/nanaimo/karst_e.php (accessed June 6, 2006)

Noga, Michael Mark. "Conference Proceedings in Geoscience Journals: What's the Use?" in *Geoscience Information Horizons: Challenges, Choices, and Decisions: Proceedings of the 38th Meeting of the Geoscience Information Society, November, 2-6, 2003, Seattle, WA, USA,* ed. Lura E. Joseph, 19-35. Alexandria, VA: Geoscience Information Society, 2004.

Northup, D.E., S. M. Barns, L. E. Yu, M. N. Spilde, R. T. Schelble, K. E. Dano, L. J. Crossey, C. A. Connolly, P. J. Boston, and C. N. Dahm. "Diverse Microbial Communities Inhabiting Ferromanganese Deposits in Lechuguilla and Spider Caves." *Environmental Microbiology* 5.11 (2003): 1071-1086.

Osif, Bonnie A. "Transportation Information: A Review of Grey Literature by Format, Language, and Availability." *International Journal on Grey Literature* 1.1 (2000): 12-17.

Palmer, Arthur N. “The Role of Cave Exploration in Karst Research.” *Journal of Cave and Karst Studies* 58.1 (1996): 4-5.

Scott, Mary W. "Status of Bibliographic Control of Pre-1900 Geoscience Literature," in *Geoscience Information Horizons: Challenges, Choices, and Decisions: Proceedings of the 38th Meeting of the Geoscience Information Society, November, 2-6, 2003, Seattle, WA, USA*, ed. Lura E. Joseph, 105-08. Alexandria, VA: Geoscience Information Society, 2004.

Straus, Lawrence Guy. “Caves: A Paleoanthropological Resource.” *World Archaeology* 10.3 (1979): 331-339.

Sulouff, Pat, et al. "Learning About Grey Literature by Interviewing Subject Librarians." *C&RL News* 66.7 (2005): 510-5.

Trivelpiece, A.W, R. S. Berry, M. Blume, J-M. Griffiths, L. Holcomb, K. McDonald, K. Rajan, K. Smith, and D. Winstanley. “Workshop Report on a Future Information Infrastructure for the Physical Sciences.” The Facts of the Matter: Finding, Understanding, and Using Information about Our Physical World. http://www.osti.gov/physicalsciences/wkshprpt.pdf (accessed July 15, 2007).

Veni, George H. “Karst and the Professional Geologist.” Paper presented to the Southwest Florida Water Management District, Tampa, Florida, 13 November 2007.

Veni, George, H. DuChene, N. C. Crawford, C. G. Groves, G. N. Huppert, E. H. Kastning, R. Olson, and B. J. Wheeler. *Living with Karst: A Fragile Foundation*. AGI Environmental Awareness Series 4. American Geological Institute: Alexandria, VA, 2001.

Walcott, Rosalind. "Guidebook Problems from a Librarian's Point of View," in *Frontiers in Geoscience Information: Proceedings of the Twenty-Fourth Meeting of the Geoscience Information Society, November 6-9, 1989, St. Louis, Missouri, USA,* ed. Mary B. Ansari, 185-92. Alexandria, VA: Geoscience Information Society, 1990.

Weintraub, Irwin. "The Role of Grey Literature in the Sciences". Brooklyn College Library. http://library.brooklyn.cuny.edu/access/greyliter.htm (accessed July 15, 2006).

Williams, P.W., ed. *Karst Terrains: Environmental Change and Human Impact*. Catena Verlag, Cremlingen-Destedt, Germany, 1993.

Chapter 13

Grey Literature Repositories: Tools for NGOs Involved in Public Health Activities in Developing Countries

June Crowe, Gail Hodge, and Daniel Redmond
Information International Associates Inc., USA

13.1 Background

Information International Associates, Inc. (IIa), a woman-owned, small business specializing in information management, performs research for government and commercial clients. IIa's Research Division has been involved in over 120 studies in the area of public health in less developed countries and regions. The information needed to complete the studies covers a range of health system topics that include statistics for health personnel, infrastructures, disaster preparedness, health financing, and other factors that impact public health care. In our experience, the search for global public health information can be both complex and frustrating. Although this information is often considered "open source" in many countries, it may be difficult to obtain, especially if governmental web sites are not readily available or completely viable, either not functioning at all or only functioning intermittently. In addition, the health information available from the site may be out of date. Many developing countries experience catastrophic events that impact access to public health information. For example, if a country has experienced political instability, natural disaster, civil strife, or other events, the existing medical system may easily be overwhelmed, with resulting health information being minimal at best.

The information used to complete these studies may also be "open access," which means the information is digital, online, available free of charge, and normally free of most copyright and licensing restrictions. The Budapest Open Access Initiative summarizes "open access" information as a wider degree of access made possible by its free availability on the public Internet, permitting any users to fully disseminate its contents without financial, legal, or technical barriers other than those needed to access the Internet itself as long as the integrity of the au-

thors' works are kept by properly acknowledging them or by using proper citations.[1]

Therefore, various resources are consulted for global public health information, including electronic journals, databases, web sites, reference sources, library catalogues, bookstores, newspapers, statistics, electronic books, maps, directories, and grey literature sources. Non-governmental organizations (NGOs) are one of the primary sources of grey literature used for researching healthcare information for developing countries.

In this publication we describe the role of NGOs in global public health information, elaborate on the problem with NGO grey literature, and describe a possible solution based on the repository concept.

13.2 Role of NGOs in Public Health Care

NGOs play an important role in global health activities and health research. It is difficult to quantify the number of such organizations. There are 53,750 development organizations listed in the 2008 edition of the Directory of Development Organizations (DDO). The DDO states that these development organizations facilitate international cooperation and knowledge sharing among civil society organizations, research institutions, governments, and the private sector.[2] According to the World Health Organization (WHO), 70-95% of health services in emergency situations are delivered by NGOs.[3] The work of many NGOs overlaps, making it difficult to discern those that have a primary focus on health. For instance, NGOs with a focus on sustainable development may also be concerned with poverty, education, and health. In Ecuador, for example, Fundacion FEVI is a non-profit NGO which facilitates intercultural education and volunteer community service. FEVI arranges community service visits from people all over the world to small communities in Ecuador. They work with healthcare centers in addition to centers for elderly people, women's organizations, indigenous communities, human rights organizations, and public schools.[4]

NGOs play key roles in health systems of developing countries and are recognized for developing innovative initiatives and programs that address health issues. They possess extensive knowledge of local conditions and can provide baseline data on health infrastructure, personnel, and major obstacles to improvement.

---

1 Suber, P. (2007), Open Access Overview. Focusing on open access to peer-reviewed research articles and their preprints. http://www.earlham.edu/~peters/fos/overview.htm

2 DEVDIR (2008), Directory of Development Organizations Home Page. http://www.devdir.org

3 World Health Organization. (2002), WHO and Civil Society Linking for Better Health. http://www.who.int/civilsociety/documents/en/CSICaseStudyE.pdf

4 Fundacion FEVI (2008), Fund for Intercultural Education and Community Volunteer Service. http://www.fevi.org/

NGOs are often able to reach segments of rural populations that governments neglect or do not target as a priority.[5]

NGOs have roles in public health from the grass roots level to the national and international levels. The WHO has created the following table depicting the health system functions and examples of roles of civil society organizations (CSO)—a type of NGO (table 1).[6]

Table 1

Health System Function	**Examples of Roles of CSOs**
Health services	Service provision; facilitating community interactions with services; distributing health resources such as condoms, bed nets, or cement for toilets; and building health worker morale and support.
Health promotion and information exchange	**Obtaining and disseminating health information; building informed public choice on health; implementing and using health research;** helping to shift social attitudes; and mobilizing and organizing for health.
Policy setting	Representing public and community interests in policy; promoting equity and pro-poor policies; negotiating public health standards and approaches; building policy consensus, **disseminating policy positions**; and enhancing public support for policies.
Resource mobilization and allocation	Financing health services; raising community preferences in resource allocation; mobilizing and organizing community co-financing of services; promoting pro-poor and equity concerns in resource allocation; and building public accountability and transparency in raising, allocating, and managing resources.
Monitoring quality of care and responsiveness	Monitoring responsiveness and quality of health services; **giving voice to marginalized groups**, promoting equity; representing patient rights in quality of care issues; and **channeling and negotiating patient complaints and claims.**

Some of these roles already involve research and information dissemination as indicated by the highlighting of those functions in the table above. Although NGOs promote and advocate for public health, as well as perform other functions in the health systems, there is a need to more effectively include NGOs in the knowledge production and diffusion of public health information in developing countries and to better manage the knowledge output.

5 Partnership with NGOs and Civil Society (2009), International Federation of Agricultural Development. http://www.ifad.org/ngo/index.htm

6 Strategic Alliance: The Role of Civil Society in Health (2001), Civil Society Initiative. External Relations and Governing Bodies. World Health Organization. Discussion Paper No. 1 CSI/2001/DP1. http://www.who.int/civilsociety/documents/en/alliances_en.pdf

13.3 NGOs and Grey Literature

Grey literature is defined as "that which is produced on all levels of government, academics, business and industry in print and electronic formats, but which is not controlled by commercial publishers."[7] NGOs create grey literature in the form of reports, online newsletters, blogs, etc. However, as mentioned above, there is a need to increase involvement of NGOs in the management of their knowledge output. This can be accomplished through dedicated partnerships with appropriate organizations and agencies. These roles could easily be expanded to include more of a role in health research knowledge diffusion because they are “on the ground” and know what is happening firsthand. A researcher having ready access to reports, online newsletters, or blogs generated by NGOs would be extremely valuable.

As a research organization, IIa and its clients need persistent access to documents from all organizations/agencies involved in health activities in developing countries. We found that for a country study completed in 2003, 18% of the urls in the study are now dead links, 3% have changed, 4% have moved or been redirected, and 29% were no longer existent. Further, the reliability for older studies becomes even more problematic. A quick look at the urls from a study completed in 2000 revealed that only 30% of the urls were active and accessible, about 62% were dead links, and about 8% of the links had been moved or had been redirected. It is widely recognized that grey literature, while frequently placed on the web only transiently, remains poorly organized and difficult to access.[8]

13.4 Repository Definition

Given the importance of NGO information and the problems mentioned with accessing this information, what could be done to improve the situation? A repository is one possible solution to the problem of locating NGO public health information, particularly reports and studies. What is a repository? A repository is a digital collection that captures and preserves the intellectual output of an institution, agency, or organization. However, it is not only the collection itself; a repository is also the services and technologies - the infrastructure - that make possible the maintenance and dissemination of the digital materials. The development of repositories has principally been undertaken by universities to collect and manage the output of students and faculty; however, they could easily be developed and used by NGOs. University development of digital repositories has been crucial in

7 The New York Academy of Medicine - Library - What is Grey Literature? http://www.nyam.org/library/pages/what_is_grey_literature

8 Liddy, Elizabeth D., Anne M. Turner, and Jana Bradley (2003), Modeling Interventions to Improve Access to Public Health Information. *AMIA ... Annual Symposium proceedings [electronic resource]* 2003: 909.

the lifespan of the technology. As of June 2009, OpenDoar, the Directory of Open Access Repositories, lists 1,407 academic repositories from around the world. One of the largest groups, approximately 7%, has a Health and Medicine subject focus.[9]

13.4.1 Institutional Repository

The focus for an institutional repository is digital collection by capturing and preserving the intellectual output of a single or multi-university community, providing a compelling response to two strategic issues facing academic institutions. This collection provides a critical component of scholarly communication, expanding access to research while maintaining control. Repositories also serve as indicators of a university's quality to demonstrate research activities and serve to increase an institution's visibility, status, and public value.[10] "University-based institutional repositories are 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 community members. It is most essentially an organizational commitment to the stewardship of these digital materials, including long-term preservation where appropriate, as well as organization and access or distribution." Institutional repositories have also been adopted by government agencies, museums, and corporations and can serve different roles in each environment. While some have argued that the primary role of institutional repositories is open access to research, others have argued that the most important function is to preserve at-risk materials like grey literature.[11]

13.4.2 Benefits of a Repository

The benefits to researchers of having one or several resources for locating and accessing this grey literature are obvious. Significant time would be saved, and there would be more assurance that the information would be updated and preserved over time. However, there are additional benefits beyond the traditional

9 OpenDOAR (2009), OpenDOAR Chart - Subjects in OpenDOAR - Worldwide. http://www.opendoar.org/onechart.php?cID=&ctID=&rtID=&clID=&lID=&potID=&rSoftWareName=&search=&groupby=cl.clTitle&orderby=cl.clCode&charttype=bar&width=600&caption=Subjects%20in%20OpenDOAR%20-%20Worldwide

10 Barton (2005), MIT Libraries. Creating an Institutional Repository. http://www.dspace.org/implement/leadirs.pdf

11 Sarah L. Shreeves and Melissa H. Cragin (2008), Introduction: Institutional Repositories: Current State and Future. *Library Trends* 57, No. 2: 89-97.

functions, such as data collection, searching, capacity building, knowledge management, as well as unified access.[12]

13.4.3 Data Collection and Coordination

An NGO repository would facilitate the identification of public health problem areas, data collection, and problem solving for decision makers. In addition to making health information about these areas more accessible to researchers and decision makers, use of the repositories could facilitate coordination among NGOs and others who want to provide assistance to these countries. A repository could be useful in identifying NGOs that have had experience in certain areas by preserving a record of the NGOs' work. It would then be easier to discern where resources could best be used.

13.4.4 Building Health Capacity in Developing Countries

Repositories could serve as a mechanism for building health capacity knowledge and diffusion in developing countries. For example, a repository could be the mechanism for introducing new perspectives, or technical expertise, and a way to capture a snapshot of what is happening with disease control, vaccinations, health education, etc. In a recent article on open access archiving, Leslie Chan pointed out that scientific progress is greatly hampered in developing countries by their inability to have access to essential medical literature.[13] A repository of NGO reports and documents could centralize access to global NGO health-related documents, particularly to those documents from other developing countries that are most relevant for public health, social, and technical situations of a developing country.

13.4.5 Knowledge Management Tool

There are direct benefits to NGOs. Those NGOs that publish many reports and documents would benefit from a repository to support content and knowledge management activities. The management of information about research and projects already conducted can support the re-purposing of that information to enhance development, marketing, and outreach efforts, as well as the creation of future funding proposals. Several years ago IIa helped Conservation International,

12 Bailey, C.W. (2008), Digital Scholarship, Institutional Repositories, Tout de Suite. http://www.digital-scholarship.org/ts/irtoutsuite.pdf

13 Chan, Leslie, B. Kirsop, and S. Arunachalam (2005),Open Access Archiving: the Fast Track to Building Research Capacity in Developing Countries. http://www.scidev.net/en/features/open-access-archiving-the-fast-track-to-building-r.html

an international environmental NGO, identify ways it could better capture and manage the knowledge created by its individual projects and principal investigators in environmental hot-spots across the globe. Development of Conservation International's system continues to this day in the ongoing implementation of a content management system for creating, disseminating, locating, and repurposing its web site content. Similar approaches would be reasonable for large public health NGOs.

A repository is a major component of an information asset management system that would manage and support every aspect of information creation and dissemination. Information asset management is the ability for people to get whatever information they need, anywhere, anytime, and in compliance with the organization's policy. As part of this function, a repository would enable the NGO to identify best practices, focus on key projects and their users, and look for partnering opportunities.

13.5 Barriers/Challenges to Repository Development

Unfortunately, there are many obstacles to the development and use of such a repository or series of repositories due to insufficient funds earmarked for health problems in developing countries, inefficient application of resources, and lack of technology transfer.[14] In this chapter, three barriers/challenges are highlighted - organizational structure and politics, funding, and collection development policies.

13.5.1 Organizational Structure and Politics

A key challenge in establishing a repository for NGOs is their wide variation in organizational structure that includes confederations, federations, separate and independent organizations, and variations of these.[15] With all these possible structures, the challenge is to create a model that will facilitate the transfer/capture of documents from all of them. Notwithstanding the fact that some NGOs do not work together due to political or philosophical differences. Authors normally deposit versions of their articles and follow a self-archiving method predetermined by the administrator's metadata policy guidelines. This process normally takes 5-10 minutes. A challenge that may arise in this type of situation would be barriers in naming convention standards and more importantly legal issues that may arise from copyrights on any formally published works. Copyright and publisher poli-

14 Delisle, Helen, et al. (2004), The Role of NGOs in Global Health Research for Development. Health Research Policy and Systems. Vol. 3(3). 2005. http://www.health-policy-systems.com/content/3/1/3

15 NGOs and Organizational Structure: Challenges and Opportunities (2003), Link no longer available.

cies of the country and/or the organization need to be considered when depositing, because normally once deposited the rights then to the publication are transferred to the repository as a whole.[16]

13.5.2 Funding

The funding source impacts how and what information an NGO releases and distributes, as well as its fiscal ability to create reports for release. For example, a religious based NGO may choose not to report on contraceptive needs or abortions, although they may have this information. Also, funding can determine which NGOs support what efforts in what countries. If several NGOs with a similar purpose, such as HIV/AIDS prevention, obtain funding from a single source, the probability of obtaining their documents for a repository is greater than if they were funded by a variety of sources, because this would perhaps eliminate some of the constraints on releasing material to the public.

NGOs may be funded by foundations, religious organizations, special interest groups, governments, international or national organizations, or any number of other methods. Their respective funding sources may impact the types and accessibility of reports or other information published. Insufficient funds, of course, may mean little or no publicly accessible information and/or the lack of a publications program. NGO funding sources can also impact the willingness to share information for political or other reasons.

In 2003, the WHO examined the funding sources of NGOs with whom they had official relationships. The majority of NGO funding (41%) came from admission fees and member dues. The next largest funding source was from unspecified grants (21%). The remainder of the funding came from other fund raising (12%); NGO grants (4%); company funding grants (3%); government and intergovernmental grants (4%); conference and publication fees (9%); and government contracts and consultancy fees (6%).[17] It should be noted that there are more NGOs that have unofficial relationships with the WHO and are thus not reported in these statistics. As civil societies have continued to increase in number, funding has increasingly come from governments (approximately USD 1 billion) and other non-governmental agencies (about USD 1 billion annually).[18]

16 Bailey C.W. (2008), Digital Scholarship, Institutional Repositories, Tout de Suite. http://www.digital-scholarship.org/ts/irtoutsuite.pdf

17 World Health Organization (2006), Some Statistics on NGOs in Relations with WHO. http://www.who.int/civilsociety/csi_statistics/en/print.html

18 World Bank Group (2005), World Bank Funding for Civil Society. http://web.archive.org/web/20050305024706/http://web.worldbank.org/WBSITE/EXTERNAL/TOPICS/CSO/0,,contentMDK:20094251~menuPK:220439~pagePK:220503~piPK:220476~theSitePK:228717,00.html

13.5.3 Collection Development Police

Given the issues outlined above, it may be impossible and perhaps not even desirable to collect all NGO documents in a single collection. A policy for collection development would need to be agreed upon even among a small group of NGOs with similar interests, such as HIV/AIDS or women's health. Another consideration would be the variation in the types of documents published by NGOs. Not all NGOs publish annual reports. Would preliminary reports or field reports with raw data be included? What about surveys or training manuals? These questions would need to be balanced by the current need for health information in the country.

13.6 Relevant Web Sites in the Public Health Domain

Despite the challenges, there are several examples of web sites that either begin to or already partially fill the role of a repository for grey literature in public health.

- United States Agency for International Development (USAID) http://www.usaid.gov/

The USAID library focuses on sustainable development with the primary mission of serving the information needs of USAID staff. USAID documents, reports, publications, and project summaries can be publicly accessed through the Development Experience System (DEXS), which has over 100,000 records with some 20,000 available for electronic download. Its purpose is primarily to strengthen USAID development projects, activities, and programs and make them publicly available. DEXS offers four major services: USAID contractors/grantees can (1) submit documents to the system, (2) search the DEXS database, (3) order documents (paper, electronic, CD), and (4) subscribe to free USAID reports via email. The DEXS submittal process is described in documentation available on the web site. Documents for submittal should include those documents which describe the planning, design, implementation, evaluation, and results of development assistance activities which are generated during the life cycle of the program or activity.

- Human Info NGO http://humaninfo.org

Uses Greenstone software. Has 35 to 40 Humanitarian CD Libraries on the Joint United Nations Program on HIV/AIDS (UNAIDS), community development, food and nutrition, health library for disasters, Rural Hygiene in Africa, Africa Collection for Transition, as well as others. About 5,000 copies of each library are distributed annually.

- World Health Organization (WHO) http://www.who.int/en

Site can be searched by country or health topic. The WHO Library and Information Networks for Knowledge (LNK) provide access to WHO-produced recorded information and to worldwide health, medical, and development information re-

sources. The Information Networks for Knowledge provides technical support to help improve the health-related information transfer structure in developing nations. The services are primarily for WHO headquarters, regions, and country offices; ministries of health and other government offices; health workers in Member States; other UN and international agencies; and diplomatic missions. The WHO library programs help regions and developing countries achieve self-sufficiency in providing information services to the health sector. The library has over 70,000 bibliographic records and 30,000 links to full text documents. Blue trunk libraries concept was developed by the library for installation in district health centers in Africa to compensate for the lack of current medical and health information. The collection of more than 100 books on medicine and public health is shipped in blue trunks fitted with two shelves. It is not known if CDs are part of this shipment. Unknown if there is a repository for NGO grey literature and/or the submittal process.

- Global Health Council http://www.globalhealth.org

World's largest membership alliance of healthcare personnel, NGOs, organizations, government agencies, and other public and private institutions. Mission is to ensure that information and resources are available to those who strive for improvement and equity in global health. Advocacy group who reports on world health problems to governments, public and private organizations, and the global health community. Publications section includes a variety of press releases, reports from NGOs and other agencies, notes from the field, annual reports of the Council, and other publications. Unknown if there is a repository and/or the submittal process, but it does have a member login/password.

- British Library of Development Studies (BLDS) http://blds.ids.as.uk/BLDS

Europe's largest library on international development at the Institute of Development Studies in Sussex. Extensive collection of government publications, NGO publications, World Bank, United Nations, World Trade Organization, and research institutes worldwide. They also have over 200 development journals that are scanned and selected articles added to the BLDS catalog. Online library catalogue can be searched at http://blds.ids.ac.uk/. Document delivery is via interlibrary loan; some items free to download. Not a repository, but a great prospect for finding NGO material.

- The New York Academy of Medicine (NYAM) http://www.nyam.org/library/

The NYAM Library's Online Catalog contains over 250,000 bibliographic records, 1,400 journals, as well as rare books and manuscripts primarily acquired since 1972. They have served the general public interested in access to health and medical information since 1878. Library services to aggregation and dissemination of "grey literature" in public health, disaster preparedness, and urban health through web-based portals. A growing repository digitization program for both web-based and at-site visiting users.

- Open Access Initiative (OAIster) http://www.oaister.org

The Open Access Initiative provides access to 21,984,755 records from 1,134 contributors. OAIster is a union catalog of digital resources. They provide access to digital resources by "harvesting" descriptive metadata (records) from numerous repositories, using OAI-PMH (the Open Archives Initiative Protocol for Metadata Harvesting). Collection focus is on digital records of any type and may include digital records with restricted access in addition to those that are freely available. Subject is not restricted to public health. Is not a repository, but is a good source for finding international public health material, including "grey literature."

Most of the web sites identified above are searchable by geographic area and have some project report summaries. Some sites are subject oriented, such as the Human Info NGO and the Global Health Council. The Human Info NGO has created repositories on CDs by subject area for distribution to developing nations and other interested parties. USAID has a growing database of health information from its partners and a defined process for the submittal of documents from NGOs to DEXS. The BLDS collects material in many subject areas and provides, via email notification, updates to the collection. However, the documents are not always in an electronic format, free, or current, though the library does make every effort to efficiently disseminate documents to people who request them.

The WHO web site has vast resources and pointers to documents, however, to our knowledge, it makes no effort to collect NGO material. The WHO library is primarily for WHO and its associated organizations. The WHO maintains relations with other international organizations and external partner NGOs. Formal relations with NGOs require that certain criteria be met. In January 2009, there were 185 NGOs that had official relations with the WHO.[19] The WHO also maintains informal working relations with other NGOs. Regional or national NGOs affiliated with international NGOs are usually charged with developing and implementing a program of collaboration with the regional and national levels of WHO in order to ensure implementation of health-for-all strategies at the country level.[20] Although WHO has the Library and Information Networks for Knowledge (LNK) that provide access to WHO-produced and recorded information as well as to worldwide health, medical, and development information resources, it has to our knowledge, neither a repository for their NGO documents nor current initiatives underway for such a repository. As can be seen in Figure 1, the number of NGO members has increased substantially over the past 19 years with 185 NGOs having formal relations with the WHO in 2009.[21]

19 World Heath Organization (2009), List of 185 Official Non-Governmental Organizations in Official Relations with. http://www.who.int/civilsociety/relations/ngolisteb120.pdf

20 World Health Organization. Relations with Other International Organizations and External Partners NGOs. http://w3.shosea.org/en/Section1257/Section1259_5127 Link no longer available.

21 World Health Organization (2009), List of 185 Official Non-Governmental Organizations in Official Relations with. http://www.who.int/civilsociety/relations/ngolisteb120.pdf

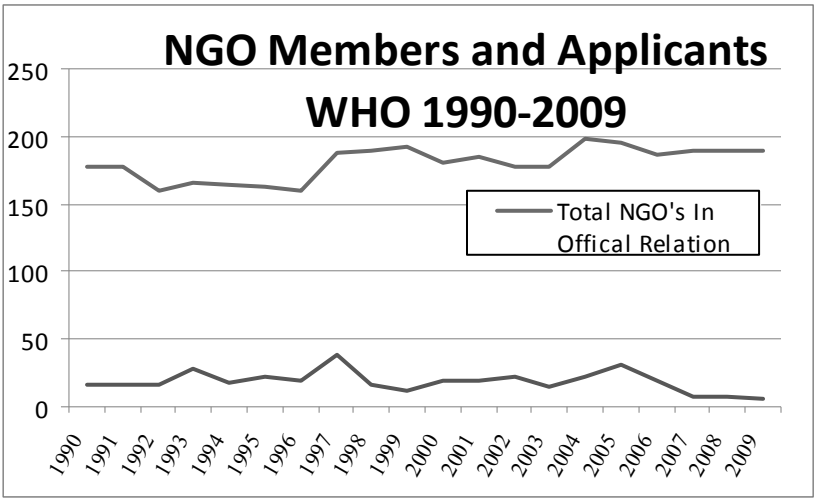

Figure 1: NGO Members and Applicants with Official Relationships with the WHO – 1990-2009.

These examples, while scattered, could serve as the basis for more consistent repository development. However, a more community-wide effort is needed to achieve this goal.

13.7 Repository Models and Platforms

Assuming that the barriers could be overcome, there are several repository models that may be viable for emulation in developing a repository and several platforms on which formal repositories by NGOs could be built.[22] These include:

13.7.1 Repository Models

PubMed Central is a digital archive of life sciences and biomedical journal literature developed and managed by the National Center for Biotechnology Information at the U.S. National Library of Medicine (NLM). This system features required participation for all investigators funded by the NIH, public release dates within one year of original publication, and retention of copyright by the author or corporate sponsor. In January 2008, the National Institute of Health's (NIH's) new policy on enhancing public access to archived publications was implemented. Authors are now required to submit an electronic version of their final manuscript

22 Johns Hopkins University (2003), Scholarly Communications Group. Publishing Models. http://openaccess.jhmi.edu/publishing.cfm

to PubMed Central upon acceptance for publication. The Policy is intended to: (1) create a stable archive of peer-reviewed research publications resulting from NIH-funded research; (2) ensure the permanent preservation of these vital, published, research findings; (3) secure a searchable compendium of these peer-reviewed, research publications that NIH and its awardees can use to manage more efficiently and to better understand their research portfolios, to monitor scientific productivity, and ultimately, help set research priorities; and (4) make published results of NIH-funded research more readily accessible to the public, healthcare providers, educators, and scientists.[23] Such a model may work for NGOs, especially if they have partners or other organizations assisting them in their work.

DSpace at MIT (Massachusetts Institute of Technology) is a digital repository created to capture, distribute, and preserve the intellectual output of MIT. DSpace features access to content through the web. Similar to PubMed Central, DSpace at MIT (and other DSpace institutions) uses the submission model; however, participation at MIT is voluntary. Authors from among the faculty provide their final manuscripts to the DSpace system. Some initial information is provided along with the manuscript, and then a "bibliographic record" or metadata file is finalized by library staff. The manuscripts are grouped into collections that represent particular communities of interest, academic colleges, or disciplines. DSpace at MIT offers the advantage of digital distribution and long-term preservation for a variety of formats, including text, audio, video, images, datasets, etc., and the opportunity to provide access to all the research of the institution through one interface.[24]

Google Scholar http://scholar.google.com/ is a search service that allows users to search for scholarly material across the web from web sites that are deemed scholarly and view either abstract or full text search results.[25] Special metadata is no longer necessary for all the pre-publication versions of papers which are deposited anywhere on the web.[26] Submission indexing eliminates the need for an NGO to develop an elaborate search system for its own documents. Much of Google Scholar's index is a subset of the larger Google search index consisting of journal articles, technical reports, preprints, theses, books, and other scholarly documents. Google scholar has built a very strong medical index, partly due to its ability to crawl full-text journals as well as specialized bibliographic databases such as PubMed.[27] Google Scholar has improved many of its features to accommodate the

23 NIH Public Access Policy (2008), Department of Health and Human Services. http://grants.nih.gov/grants/guide/notice-files/NOT-OD-08-033.html

24 Johns Hopkins University (2003), Scholarly Communications Group. Publishing Models. http://openaccess.jhmi.edu/publishing.cfm

25 Sullivan, D. (2004), Google Scholar Offers Access to Academic Information. Search Engine Watch. http://searchenginewatch.com/searchday/article.php/3437471

26 ALPSP (2005), Preprint and postprint repositories and their impact on publishing. http://www.keyperspectives.co.uk/openaccessarchive/Conference%20presentations/Swan%20-%20ALPSP%20IR%202005.pdf Link no longer available.

27 Vine. R. (2006), Journal of the Medical Library Association, Google Scholar. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1324783

demands for medical searches. The advanced search has recently gone live, which provides subject area searches, author searches, and, most importantly, the ability to return articles published within date ranges. Google Scholar's unique ranking feature then ranks the returned results with the most relevant results appearing first.

Open Access Archives (OAAs) is another model that would encompass the variety of types of information published by NGOs. Open Access Archives are repositories that have a policy of providing journal articles free and online. Typically this is done using self-archiving; the NGO would submit documents to an institutional or community archive of its choice such as arXiv.org, CiteSeer, or another repository that was appropriate for the content. This model is favored in the article by Leslie Chan as a quick way to build research capacity in developing nations.[28]

For small NGOs, the approach of the **Association of Learned and Professional Society Publishers (ALPSP)**, http://www.alpsp.org/ngen_public/, may be of interest. These are "community" organizations that have been created to build the capacity of the small publishers using the digital environment.

13.7.2 Repository Platforms

Digital Commons, hosted by the Berkeley Electronic Press (bepress), is the largest manufacturer hosted repository platform that can help institutions collect, showcase, and preserve scholarly output. They build the repository to match the institution's web site and provide unlimited technical support. Digital Commons features online submissions, content management, advanced indexing, support for multiple content types, and conversion of popular document formats to PDF. Digital Commons offers its customers a platform for repository development that guarantees 99.9% uptime, unlimited tech support, and setup of the system in 1 to 2 weeks.[29] Examples of health repositories currently using bepress software include:

Government of South Australia Department of Health http://www.publications.health.sa.gov.au/ (Australia);
Houston Academy of Medicine, Texas Medical Center http://digitalcommons.library.tmc.edu/ (USA);
Royal College of Surgeons Ireland http://epubs.rcsi.ie/ (Ireland)

DSpace has been a pioneer in open source digital repository software and is the most commonly used software platform for developing an institutional repository. DSpace is not a hosted solution, but their site does provide links to numerous service providers if an institution does not have the technical expertise or re-

28 Chan, Leslie, B. Kirsop, and S. Arunachalam (2005), Open Access Archiving: the Fast Track to Building Research Capacity in Developing Countries. http://www.scidev.net/en/features/open-access-archiving-the-fast-track-to-building-r.html

29 Bepress, The Berkeley Electronic Press Home Page. http://www.bepress.com/index.html

sources for developing a repository. DSpace is free open-source software, released under a BSD license, that is easy to implement and completely customizable. DSpace supports a wide variety of formats and features a large user community and discussion forums for obtaining technical assistance.[30] Examples of health repositories currently utilizing Dspace software include:

College of Public Health Sciences http://cphs.healthrepository.org/ (Thailand);
University of Calgary E-Health Repository https://dspace.ucalgary.ca/handle/1880/42949 (Canada);
WHO EHA Institutional Repository http://whoindonesia-eha.healthrepository.org/ (South East Asia)

EPrints is a UK-based open source software flexible platform for building high quality and high value repositories. It is the self-proclaimed easiest and fastest way to set up repositories for research output from literature, scientific data, and reports through archived documents, multimedia, or documents. According to EPrints own database of repositories, there are currently 269 known implementations of EPrints repositories, which are mostly found in Europe. However, the Registry of Open Access Repositories (ROAR) lists 333 known repositories at the time this chapter was written. The EPrints Services team offers fee-based advice and consultation that ranges from initial help all the way through to a completely managed institutional repository.[31] Examples of health repositories currently utilizing EPrints software include:

University of Birmingham School of Health Sciences: http://eprints.bham.ac.uk/view/divisions/sch_heal.html (UK);
University of Nottingham Department and Faculty of Medicine and Health Sciences: School of Clinical Laboratory Sciences http://eprints.nottingham.ac.uk/ (UK)

Fedora Repository Project is an architecture for developing an institutional repository system. The current community project has been released as the Fedora Repository Project and the community responsible has been officially named the Fedora Commons. The current (2009) release of Fedora Repository offers advanced database technology for digital content preservation and advanced features such as messaging (for within site help) and administrative clients. Fedora has been growing very rapidly in popularity due to its strong technology, excellent data handling, and very active community. Since it is open source software, institutions also see the high benefit of not having to pay licensing fees.[32] The most notable example of is from the University of Prince Edward Island Robertson Library. It is commonly referred to as “Icelandora” within the development community. http://library.upei.ca/ ;

30 DSpace Home Page. http://www.dspace.org
31 Eprints, Eprints Home Page. http://www.eprints.org/
32 Fedora Commons Software Home Page. http://www.fedora-commons.org/

Another example is The Australian Research Repositories Online to the World http://arrow.edu.au/ (Australia)

Greenstone digital library software is free, open-source, and multi-lingual platform for developing a repository and publishing it on the Internet or on CD-ROM. An NGO could use this software to build its own digital libraries. Greenstone is produced by the New Zealand Digital Library Project at the University of Waikato and developed and distributed in cooperation with the United Nations Educational, Scientific and Cultural Organization (UNESCO) and the Human Info NGO. The Human Info NGO is based in Antwerp, Belgium and works with United Nations agencies and other NGOs. They have established a worldwide reputation for digitizing documents in human development and making them widely available and free to developing nations and on a cost-recovery basis to others. A new development with Greenstone is the ability to build collections on a remote server while using a modified version of the Greenstone Librarian Interface, so there is no need to run Greenstone locally. Multiple users can collaborate on the same collection, although not simultaneously.[33]

The software for the basic development of a repository is available, and most of it is open source. Greenstone has the additional benefit of being multi-lingual and portable. However, the submission and/or harvesting approaches for capturing grey literature must be carefully considered, as would a collection development policy.

13.8 Conclusions

As an information management and research company, IIa believes that grey literature is a vital component of public health information, particularly in developing countries. One or more repositories of grey literature from across NGOs in the public health community would be beneficial to researchers seeking to use this information. While there are many barriers to achieving such a repository, the benefits would be numerous and a variety of models could be used. There are several existing web sites that begin to fill this need, but a more community-wide effort is required in order to provide consistent, complete, and effective coverage of this grey literature. While the benefits to the research community are obvious, the ultimate benefit is to advance the use of public health research in improving the lives of people world-wide.

33 Greenstone Digital Library Software. Home Page (2009). http://www.greenstone.org/

Part II, Section Five

Future Trends in Grey Literature

What does the future hold in store for grey literature? Some people anticipate that since Internet and Google push an unimaginable amount of information to the user, grey dissemination channels will eventually disappear. However, this notion is not shared by all.

As we stated in our Introductory Chapter, grey literature will not disappear, but will instead continue to play a significant role alongside commercial publishing even if the borderline between "grey" and "white" (commercial) literature will become increasingly indistinct. This holds particularly true in an environment shifting towards open access to scientific and technical information. Actually, we expect that the proportion of "grey" documents published on the Web will continue to increase and the Internet will instead encourage a greater diversity in the types of "grey" resources available, such as raw data, personal notes, lectures, etc.

Our predictions are based on empirical data and observations, which show that despite the rapid development of the open access movement only a part of reports and theses have become freely and easily available on the web. While other types of grey documents remain virtually inaccessible. Thus, in this section we choose not to focus on whether grey literature has a future or not, but instead through the eyes of four information professionals, we examine new environments of mediation and information transfer as well as innovative perspectives for non-commercial documents and their dissemination.

The first chapter in this section begins by looking to new forms of scientific communication. For Banks, "findability" of grey literature "is a less pressing concern than before". Banks turns to consider the preservation of Web2.0 content, particularly from blogs and twitter. Based on discussion and a case study, he urges that "general digital preservation principles combined with an evolving understanding of the uses of Twitter would be necessary in developing preservation criteria for blogs and tweets."

In the previous section, we saw how Gentil-Beccot appeals for increased investment in open archives, especially institutional repositories. However, we are still left with the measure of return on investment? The second chapter in this section by Schöpfel and Boukacem confronts some of the financial aspects of grey literature in institutional repositories (IR). "Grey does not mean free." Until now, the problem has been that little is known about repository costs and usage statistics. This chapter attempts a state of the art and suggests some COUNTER derived

metrics that may assist in comparing archives and their investment policies, such as IR costs per item, IR costs per user, and IR items per scientific output.

The third chapter in this section by Jeffery and Asserson places grey literature in the context of eScience. The authors introduce the e-research environment and describe the European CERIF format for current research information systems (CRIS), which allows for interoperability between systems and institutions. Their reasoning is twofold: first, grey literature should be stored in open repositories and second, the metadata should be compliant with the CERIF format and stored in the current research information system. They explain that "with the two sources linked to allow optimal use of the characteristics of the CRIS and the repository (…) not only is the grey literature object provided with better metadata for retrieval but also is associated with the other contextual metadata in the CRIS". This would include projects, persons, organisations, facilities, equipment, events, products, patents, etc. They continue "this further places grey firmly in the research environment together with other publications and products. This architectural approach positions optimally grey literature."

The fourth and final chapter in this section provides an overview on grey literature in higher education - not as a resource but as an object of teaching in order to "gauge the current place of grey literature in library and information science education". Rabina examines here course descriptions and syllabi among the 2009 top ten LIS graduate programs in the United States. She concludes by recommending grey literature be taught in cross-curricular programs in accordance with the interdisciplinary scope of grey literature content.

In comparison with the other sections in this book, this section remains quite open-ended. While there is no one final conclusion, we do ask the reader to bear in mind a few key questions:

Should grey literature be linked to primary research data (datasets) and if so, how? How can the quality of grey items be assessed and guaranteed? Do usage patterns differ between grey items and journal articles, books, etc.? How should the concept of grey literature be adapted to the emerging environment of eScience? How can LIS schools and colleges adequately ensure the coverage of grey literature in their curricula programs? And, last but not least, what kind of empirical evidence should be produced in order to develop a better understanding of non-commercial scientific information?

Chapter 14

Blog Posts and Tweets: The Next Frontier for Grey Literature

Marcus Banks, UCSF Library, USA

14.1 Introduction

My interest in grey literature began as an Associate Fellow of the US National Library of Medicine (NLM) from 2002-2004. Many colleagues at NLM and throughout the country work to improve information access for people in the public health workforce. In comparison to resources available for practicing clinical medicine, the information needs for public health are more diffuse and often require access to grey literature [1]. In response to these needs NLM has developed the Partners in Information Access for the Public Health Workforce portal, which includes some avenues to grey literature but mostly useful links and validated search strategies for PubMed [2]. In addition to NLM's Partners page, staff of the New York Academy of Medicine Library has maintained the Grey Literature Report for several years [3]. This is a portal to documents produced by reputable organizations in public health and health policy.

During my NLM fellowship years, debate about open access publishing - specifically, how to secure access to publicly funded research - was heating up. An open access publication is generally a type of white literature, and is available for free online and stored in a digital repository [4]. Debate over the proper balance between open and subscription access will continue for years, as library associations and publishers continue to hire lobbyists and issue strongly worded statements. In 2004 I was certain that pure open access would prevail, but now think a hybrid subscription-OA model is much more likely to endure. During this more optimistic phase, I argued that grey literature advocates could learn from the political strength of open access advocates, and mount a similar campaign to demonstrate the value of grey literature [5]. Grey literature is almost always free to read already, so it only needed to be found.

Today I am much less concerned about findability for grey literature. While it remains simpler to locate a journal article than a working paper, smart Google searches can easily unearth the latter (I'll provide some examples of this in the next section.) Portals to grey literature remain useful for providing context and browsability; along with the Grey Literature Report, OpenSIGLE also serves both

these functions well [6]. Even so, the core challenge of finding grey literature in the first place is much less potent than in years past. So rather than mounting a political campaign to raise the profile of grey literature, I believe that grey literature advocates should now concern themselves with strategies for preserving the ephemeral "grey data" represented in content such as blog posts and tweets [7]. After addressing improved findability, I will present the case for my position.

14.2 Improved Findability for Grey Literature

Traditional barriers to locating grey literature, in comparison to white literature, include irregular publication schedules and the lack of standard bibliographic identifiers such as volumes, issues, and page numbers. These difficulties persist, but are much less fatal in the Internet age than they were in the print-only era.

It is possible to search Google for specific file types, and/or to restrict the search to the domains of organizations that produce a significant quantity of grey literature. Examples of each type of search, with screen shots, are found below (screen shots are current as of September 5, 2009.)

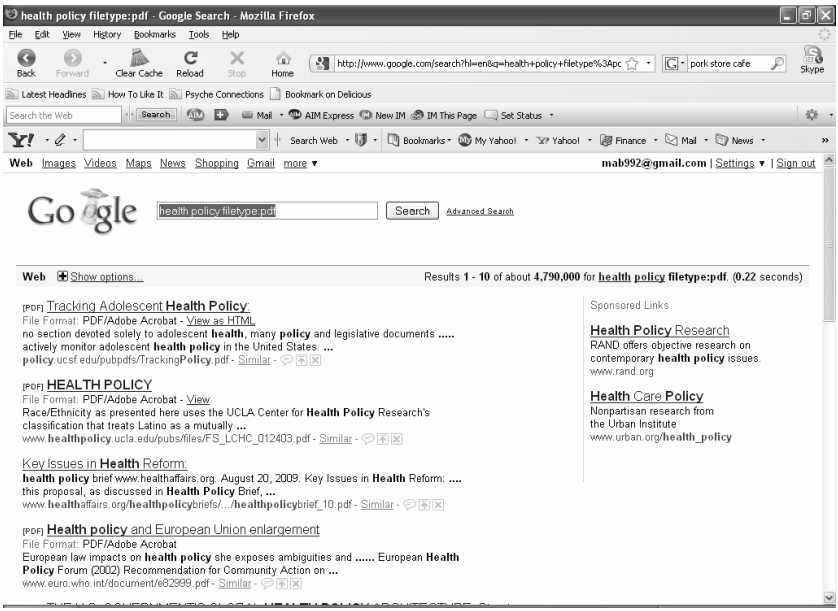

Figure 1: Search for specific file type: "health policy filetype:pdf"

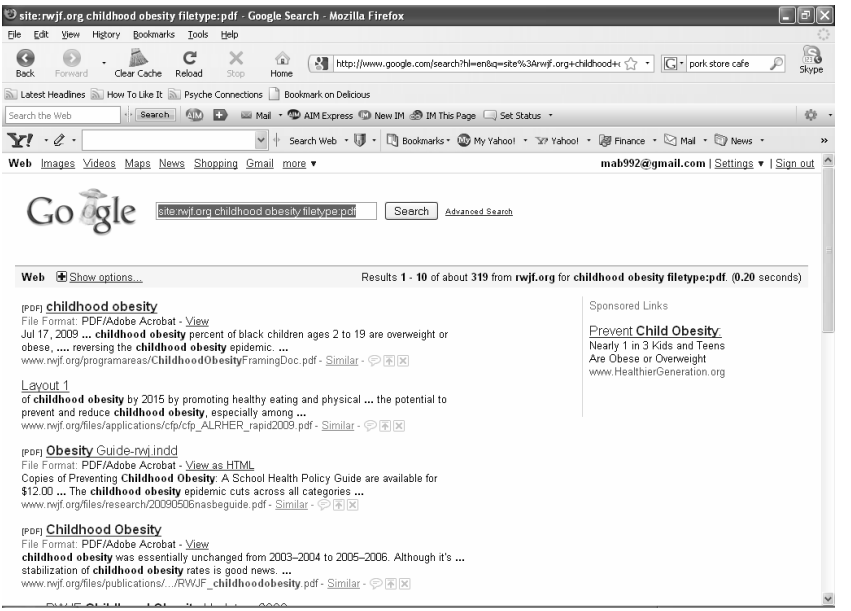

Figure 2: Search within domain of specific organization: site:rwjf.org childhood obesity filetype:pdf (Restricts to searches of the Robert Wood Johnson Foundation domain)

These examples demonstrate how easy it is to locate grey literature, *as long as searchers know the appropriate search strings*. I recognize that this syntax is obscure to a very large majority of end users. However, librarians can instruct patrons about these search strategies, and also develop search systems that invoke these kinds of strategies behind the scenes.

Findability is not the only concern with grey literature. In the first example above the documents come from a large smattering of organizations, and would be of varying use depending on an individual's information needs. In the second search, I needed to know in advance that the Robert Wood Johnson Foundation was a useful source of information. The advantage of OpenSIGLE and the Grey Literature Report, in comparison to "raw Google," is that they both organize materials and vet their sources.

That said, findability *per se* is no longer a concern for traditional grey literature.

14.3 "Web 2.0" Content as "Grey Data"

If we accept the premise that findability is a less pressing concern than before, how should the grey literature community focus its energies today? I submit there is a pressing need to preserve the content being generated via various Web 2.0 tools and platforms.

First, some background: the somewhat annoying moniker "Web 2.0" describes the more interactive Internet that has emerged in recent years [8]. First generation web sites tended to be static HTML pages where readers could look but not touch, and the only way to share anything was to forward web links via email. Today it is extremely easy to post articles or news clips (say, from the *New York Times* or the BBC) to one's Facebook page. (Facebook is just one of many "social networking" services; others include LinkedIn or MySpace.) This enhances sharing between friends and colleagues, which is useful. But the most profound change lies in the ability for anyone to post "user-generated content" such as blog posts or YouTube clips. Web 2.0 tools are also beginning to influence scientific debate [9].

Blogs are now an established part of the information landscape; they are scrolling public diaries that usually allow comments. Twitter is a more recent development. It is a "micro-blogging" service that enables users to post very short messages (no longer than 140 characters) via their web browsers or mobile phones. Each message is a "tweet," and "retweeting" interesting messages has emerged as a rapid way to broadcast information [10].

I have maintained a blog—which contains ruminations on both professional and personal matters—since January 2005 [11]. Many of my colleagues in health sciences libraries also write blogs, often with a mix of personal and professional content [12, 13, 14]. In 2004 I did not read any blogs for professional information, and today they are critical professional sources. Blogs are updated much more rapidly than traditional journals, and readily facilitate conversation (although people often do not comment on posts.) Many librarians now tweet, sometimes to make pithy observations but often to share interesting blog posts or other online content [15, 16]. Within Facebook, the status update is functionally similar to a tweet.

The various Web 2.0 tools have enhanced both my professional and personal life. They enable a more fluid and informal form of communication, and in some form they are here to stay. (This Web 2.0 stuff could all be a fad, but at least within the library realm I think it's unlikely that we'll return to journal articles and white papers as dominant distribution mechanisms.) If this premise is correct, then a major concern about reliance on Web 2.0 tools is that we do not yet have good mechanisms for permanently archiving content produced with these tools.

Tweets are almost by definition ephemeral, and blog posts suffer from the general "link rot" that bedevils the Web [17]. Commercial solutions for Twitter archiving are emerging [18], and at least we have the Internet Archive [19]. But we are not yet close to an equivalent to acid-free paper for content developed online. Grey literature advocates can step into this breach, particularly if we expand the definition of grey literature to include the more informal "grey data" [20].

In the meantime, there are ad hoc efforts to preserve Web 2.0 content. For example, in the summer of 2009 the US Library of Congress announced plans to preserve tweets associated with Justice Sonia Sotomayor's successful nomination to the Supreme Court [21]. An endorsement of the value of tweets (and by exten-

sion of blog posts, as many of the tweets referred to blogs commenting on the nomination) by an institution such as LC is a powerful indicator of their importance. Given LC's decision, to close this chapter I will provide further arguments for why the grey literature community should accept the challenge of preserving Web 2.0 content.

14.4 Case Study: Nicole Dettmar and Clinical Reader

Before making the argument for preservation of Web 2.0 content as grey literature or "data," let's examine the case of my librarian colleague Nicole Dettmar. Her experience in the summer of 2009 points to the vital need for permanent archiving of tweets in order to understand how people communicate online today.

Dettmar blogs, and in July 2009 she skillfully criticized the web site Clinical Reader for falsely implying that it had earned endorsements from leading libraries and for using copyrighted images without permission [22]. The critique was both accurate and thoughtful. Clinical Reader's initial response, via Twitter, was to "kindly request" that Dettmar remove her blog post or else face the risk of legal action [23]. As of September, Clinical Reader has apologized to Dettmar and removed the implication of non-existent endorsements.

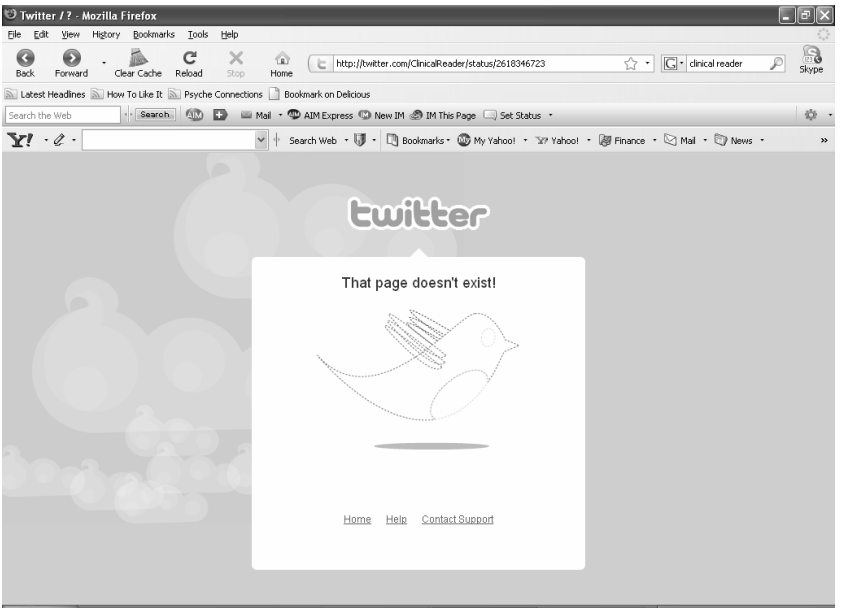

Figure 3.

In the wake of Clinical Reader's overwrought response to her post many librarians, as well as *Guardian* columnist Ben Goldacre, leapt to her defense. Much of

the conversation between Clinical Reader and its critics took place on Twitter. But it was a hard conversation to have, because representatives from Clinical Reader consistently deleted their tweets and re-emerged with new Twitter accounts. Dettmar utilized screen shots to preserve the tweets for posterity [24], and the Disruptive Library Technology Jester also tracked the action [25]. Despite these laudable efforts, it would now be difficult for even the most intrepid scholar to piece together what happened on Twitter because of the deliberate interruptions to the flow of conversation. Below is a screen shot of a Clinical Reader Twitter screen that no longer exists, because the content was deliberately deleted.

If there is anyone who should care passionately about the preservation of otherwise overlooked and forgotten discourse, it is the grey literature community.

As Dettmar modestly stated, her initial post about Clinical Reader "was not of massive importance" [26]. But the support she received was gratifying and invigorating. Another important dimension of this support was its technical fluency. Dettmar's supporters utilized Twitter with ease, responding to her critics within Twitter and retweeting in order to bring more attention to aspects of the discussion [27]. All of this happened in real-time, at a much faster pace of discourse than existed before the Web, or even in the Web 1.0 days. Tweets are early warning devices, and the blog posts or news articles they reference provide the context for whatever controversy is brewing.

Whether or not one has much interest in the Clinical Reader controversy, the phenomenon of the "rapid stream" of comments that registered the controversy should be of interest to the grey literature community. As I write, United States residents and political leaders are debating whether and how to reform the health care system. Here too Twitter offers a vital register of the discussion. And just as with the Clinical Reader tweets, there is no guarantee that this record will persist. Unless, that is, members of the grey literature and broader information communities resolve to preserve this record and others like it.

14.5 Connection Between Web 2.0 Content and Grey Literature/Grey Data

My initial interests in grey literature stemmed from admiration at the political savvy of open access advocates, with hopes that the GL community could learn from the open access movement and raise the profile of grey literature [28]. Now that findability for grey lit is not as large a concern, my interests have shifted to the idea of a "continuum" that will eventually collapse the distinction between grey and white literature [29].

I once assumed that this continuum only included materials that could easily be printed. My idea was that peer review could often happen online rather than behind the scenes, and I conceived of blog posts as the most radical extension of traditional forms of communication. Everything I envisioned would properly fall under the heading of "literature."

As tweets can be no longer than 140 characters, it is a stretch to call them literature. But they are definitely useful bits of data, which collectively can aggregate into an important lens of understanding an ongoing discussion. In 2009 the lonely tweet suffers from the same findability problems as the working paper did pre-Google. Twitter facilitates searching for groups of tweets via the "hash tag" convention (example: #Obama), but the individual tweets buried within conversations can be easily lost [30].

If we conceive of the tweets as "grey data," then the preservation imperative for the grey literature community becomes clearer. Robust archiving services for Twitter would have prevented Clinical Reader from manipulating the Twitter conversation this year.

Blog posts leave a larger footprint than tweets, although they still ephemeral and subject to the usual Internet link rot. They are also "grey" in the sense of not containing standard bibliographic identifiers. Although tweets are particularly at risk because they are so easy to proliferate, blog posts would also be worthy of the preservation attention of the grey literature community.

This begs a whole host of questions. With literally millions of bloggers of Twitterers out there, how can anyone possibly determine what to preserve? The Library of Congress's decision regarding the Sotomayor tweets was an easy call, but this won't always be the case. While I recognize the enormity of the challenge, the developers of the New York Academy of Medicine and OpenSIGLE portals also had to establish selection criteria (albeit on a much smaller scale). General digital preservation principles, combined with an evolving understanding of the uses of Twitter, would be necessary in developing preservation criteria for blogs and tweets. I have no words of wisdom in this regard, except to say that this is clearly an area of growth for the grey literature community and that I would be happy to be part of any discussions in these areas.

14.6 Conclusion

My interests in grey literature have shifted considerably in the last five years, which is roughly synonymous with the emergence of "Web 2.0" tools. I almost feel like a fraud writing this chapter, because my interests are so divergent from what I presented at conferences in Nancy and New Orleans! But at the risk of seeming fraudulent, I really do believe that preservation of Web 2.0 content should be a main focus for the GL community in future years. Traditional grey literature remains important, but thankfully it is much easier to find than before. Let us now turn our attention to new and exciting challenges.

References

1. Alpi K.M. (2005), Expert searching in public health. – In: J Med Libr Assoc, 93 (1) Jan 2005, pp. 97-103. http://www.pubmedcentral.nih.gov/picrender.fcgi?artid=545128&blobtype=pdf
2. NLM, Partners in Information Access for the Public Health Workforce, http://phpartners.org/. [cited 7 Sep 2009].
3. NYAM Library, Grey Literature Report, http://www.nyam.org/library/greyreport.shtml. [cited 7 Sep 2009].
4. Suber P. (2003), Bethesda statement on open access publishing, http://www.earlham.edu/~peters/fos/bethesda.htm. [cited 7 Sep 2009].
5. Banks M. (2004), Connections between open access publishing and access to gray literature. – In: J Med Libr Assoc, 92 (2) April 2004, pp. 164-166. http://www.pubmedcentral.nih.gov/picrender.fcgi?artid=385294&action=stream&blobtype=pdf [cited 7 Sep 2009].
6. INIST, OpenSIGLE, http://opensigle.inist.fr/. [cited 7 Sep 2009].
7. Banks M. (2009), Open access, grey literature, grey data, http://mbanks.typepad.com/my_weblog/2009/07/open-access-grey-literature-grey-data.html [cited 7 Sep 2009].
8. O'Reilly T. (2005), What is Web 2.0?, http://oreilly.com/web2/archive/what-is-web-20.html. [cited 7 Sep 2009].
9. Economist (2008), User-generated science, http://www.economist.com/sciencetechnology/displayStory.cfm?STORY_ID=12253189 [cited 7 Sep 2009].
10. Blogging Bits (2008), How to retweet: a simple guide, http://bloggingbits.com/the-art-and-science-of-retweeting-for-twitteraholics/. [cited 7 Sep 2009].
11. Banks M., Marcus' World, http://mbanks.typepad.com/ [cited 7 Sep 2009].
12. Kraft M., Krafty Librarian, http://kraftylibrarian.com/ [cited 7 Sep 2009].
13. Plutchak T.S., T. Scott, http://tscott.typepad.com/. [cited 7 Sep 2009].
14. Rothman D., Exploring medical librarianship and web geekery, http://davidrothman.net/. [cited 7 Sep 2009].
15. Rumsey E., Eric Rumsey on Twitter, http://twitter.com/ericrumsey. [cited 7 Sep 2009].
16. Walden R., Rachel Walden on Twitter, http://twitter.com/Rachel_w. [cited 7 Sep 2009].
17. Wikipedia (2009), Link rot, http://en.wikipedia.org/wiki/Link_rot. [cited 7 Sep 2009].
18. QuoteURL (2009), http://www.quoteurl.com/. [cited 7 Sep 2009].
19. Internet Archive (2009), http://www.archive.org/index.php. [cited 7 Sep 2009].
20. Banks M. (2009), Open access, grey literature, grey data, http://mbanks.typepad.com/my_weblog/2009/07/open-access-grey-literature-grey-data.html. [cited 7 Sep 2009].
21. Library of Congress (2009), LOC will capture tweets on #sotomayor for its web archives on the Sotomayor nomination, http://twitter.com/librarycongress/statuses/2222610835. [cited 7 Sep 2009].
22. Dettmar N. (2009), Clinical Reader: Starry ethics fail, http://eagledawg.blogspot.com/2009/07/clinical-reader-starry-ethics-fail.html [cited 7 Sep 2009].
23. Lawson S. (2009), Clinical Reader: From zero to negative sixty with one bogus threat, http://stevelawson.name/seealso/archives/2009/07/clinical_reader_from_zero_to_negative_sixty_with_one_bogus_threat.html [cited 7 Sep 2009].
24. Dettmar N. (2009), Twitter search on "legal ramifications," July 16, '09 on Twitpic, http://www.twitpic.com/al4eg [cited 7 Sep 2009].

25. Disruptive Library Technology Jester (2009), On the pitfalls of social media: The case of Clinical Reader, http://dltj.org/article/clinical-reader-background/ [cited 7 Sep 2009].
26. Dettmar N. (2009), Gratitude, http://eagledawg.blogspot.com/2009/07/gratitude.html [cited 7 Sep 2009].
27. Lozeau P.M. (2009), RT @adr Dear @ClinicalReader: Stop threatening bloggers with bogus legal claims. OK? Thanks. (i.e http://ur1.ca/78oc). [cited 7 Sep 2009].
28. Banks M. (2004), Connections between open access publishing and access to gray literature. – In: J Med Libr Assoc, 92 (2) April 2004, pp. 164-166. http://www.pubmedcentral.nih.gov/picrender.fcgi?artid=385294&action=stream&blobtype=pdf [cited 7 Sep 2009].
29. Banks M. A. (2005), Towards a continuum of scholarship: the eventual collapse of the distinction between grey and non-grey literature. – In: GL7 : Seventh International Conference on Grey Literature, Nancy (France), December 2005. http://eprints.rclis.org/5803/ [cited 7 Sep 2009].
30. Tech for Luddites (2009), The Twitter hash tag: what is it and how do you use it?, http://www.techforluddites.com/2009/02/the-twitter-hash-tag-what-is-it-and-how-do-you-use-it.html [cited 7 Sep 2009].

Chapter 15

Assessing the Return on Investments in Grey Literature for Institutional Repositories

Joachim Schöpfel and Chérifa Boukacem-Zeghmouri
University of Lille, France

15.1 Good (and bad) reasons for assessment

A main feature of usage statistics is their benefits for an evaluation of return on investment, especially in the era of big deals between academic libraries, consortia and publishers. Libraries and funding organisations invest heavily and increasingly significant amounts of money in e-journals, e-books, databases and other online resources, and they need to know what they get in return – not (only) in terms of content, but in terms of value for end users.

The evaluation of value for money in the use of public spending is on the agenda of academic and research organisations. The new public research policy requires funding to be linked to performance and commitment of results. Library and information science (LIS) professionals have to justify their investment choices, and they need to show return on investment (ROI) to their resource allocators. In other terms, they must merge elements of cost analysis and usage assessment.

One explanation is that the importance of the role of the library as a gateway for locating and accessing information has fallen over time (Housewright & Schonfeld, 2008). As Lauridsen (2009) observed recently, while library expenditures keep going up, growth in usage statistics slows down.

Nobody can reasonably expect academic libraries to generate net income. But this value gap (Tenopir, 2009) asks for monitoring. Any information service needs some kind of assessment so as to improve quality and performance and to optimize the impact of public spending. "Methods of cost-benefit analysis, such as ROI, are important tools in assisting one in making informed decisions (…) and to gain more credibility from various stakeholders" (Linn, 2009).

Academic libraries look back on a longstanding tradition of statistics and metrics, and international standards facilitate assessment and comparison (ISO, IFLA; see Heaney, 2009). In spite of this tradition, the rapid development of digital resources, open access and e-science appears to challenge the LIS professionals' capacity of monitoring and assessment.

15.2 Grey business?

This chapter is about money. Not the money one can earn by providing information services. But the money public institutions spend on the acquisition, promotion, dissemination and preservation of scientific grey literature through open archives, in particular institutional repositories.

Introducing economics to grey literature may seem paradoxical because of the non-commercial character of grey literature. Compared to the academic journal market, there are only (very) few studies on business models and the value chain of grey literature (see Roosendaal in this book). Grey is often (mis)understood as free.

Of course, this is wrong. As wrong as the idea that most grey stuff one day will be published and disseminated through the usual (e.g. commercial) distribution channels. In fact, only a small part (probably not more than one third) crosses the border and becomes white – Ph.D. dissertations edited by book publishers, conference proceedings published in special issues, scientific reports edited in a serial collection. The other material never enters the information market.

One corollary of this situation is that the processing and preservation of grey scientific literature is mainly if not exclusively non-for profit business, managed by public information services on a local, national or international level.

The grey acquisition budget appears generally to be relatively low. Partly grey literature is collected without any direct expenditure, through legal deposit of research reports or submission of theses and dissertations. Yet, a grey collection bears at least indirect costs. Human resources are needed and have to be paid; other cost centres are the information system, storage facilities, records production and management, dissemination of copies, and so on.

Grey does not mean free. Correlated to the overall number of items, the acquisition of grey material may come out as more expensive than expected. Big deals with commercial publishers or database producers may be very expensive, but divided through the overall number of articles, issues or records, the item price often is rather low. On the other side, while a library may spend only a small part of its budget on grey literature, divided through the number of grey items, the individual acquisition and processing costs may be rather high.

This may seem a paradox. In fact, it highlights the value and relevance of grey literature. The important scientific and technical information (STI) centres have a specific "grey footprint" as the different chapters of this book and our own studies clearly show (Boukacem-Zeghmouri & Schöpfel, 2006; Schöpfel & Prost, 2009). They define a specific grey acquisition and collection policy, they invest in a specific way, and they offer specific services to their communities and customers.

But while some for-profit companies developed "grey" added value services such as alert products based on data mining of conference announcements and abstracts, public STI centres rather granted open (free) access to grey literature.

15.3 Grey content in institutional repositories

For political and financial reasons, STI centres are part of scientific communities and endorse their decisions. Since 2002 (OAI Budapest Declaration), universities, research organisations and scientific communities opt for and invest in the creation of institutional repositories in order to facilitate and speed up direct scientific communication and to develop an alternative to the commercial scientific information market ("serials crisis").

Following Jones (2007), an institutional repository is a safe place to store a critical mass of intellectual work in digital format, where the collection is linked to a specific organisation or community, together with (in particular) descriptive metadata and a method of finding it again. It fulfils two requirements: a method of disseminating outputs under the aegis of the organisation (outward facing), and a central location and focus for the collection of the outputs of the organisation.

For a STI centre or an academic library, the project of an institutional repository with facilities for deposit and metadata creation by the author may also in the long run simplify and rationalize the preservation, processing and dissemination especially of grey documents.

Institutional repositories are a key element of the emerging landscape of open access to research and scholarship (Willinsky, 2006). Generally considered as the "green road" to open access (Harnad et al., 2008), the number of open archives referenced by the international directory OpenDOAR increased steadily since 2007 at an annual rate of around 30% and attains today more than 1,500 sites; more than 80% are institutional repositories hosted by universities or other scientific structures. Yet, these figures underlie the reality, as surveys from Spain and France prove (Melero et al., 2009; Schöpfel et al., 2009). In France the number of open archives nearly tripled last year, growing from 56 in 2008 to 150[1] in 2009.

The part of grey literature in these archives is extremely variable, varying from 0 to 100%. Let's look at some figures:

All institutional repositories contain one or more types of grey material – often electronic theses and dissertations, but also unpublished working papers, courseware, conference proceedings or project reports.

Grey material accounts for 16% of the open archives' content in France and 21% in Spain. Nevertheless, the part of grey material is significantly higher in institutional repositories than in other categories:

Table 1: Part of grey literature in French open archives (2009)

Type of repository	Part of grey literature
Institutional	41%
Non-institutional	9%
All	16%

1 Only 52 of them are listed in the OpenDOAR directory (February 2010).

What is the relative part of the main types of grey literature? Most of the grey items in French institutional repositories (IR) are communications:

Table 2: Part of grey document in French institutional repositories (2009) (*conference proceedings; **electronic theses and dissertations)

Type of documents	Part in IR
Communications*	55%
ETDs**	19%
Reports	10%
Working papers	3%
Courseware	0,1%
Other	13%

The problem with these figures is that they depend on the definition of grey literature and also, on the repositories' metadata quality. Sometimes it is difficult to distinguish different document categories. Many repositories simply don't define their categories and probably leave it to the authors (and visitors) to make the choice. Together with the often more or less poor search facilities in repositories, the lack of standards and shared understanding makes assessment and evaluation difficult.

15.4 Usage assessment

One way to cope with the need to assess the return on investment (ROI) is the collection and evaluation of usage statistics. Projects like COUNTER and SUSHI are designed to assist publishers, vendors and libraries in this task, through the precise definition of terms and concepts, through standardization of procedures, figures and presentations, and through labelling of products (Shepherd, 2005).

The real use of individual items, journal titles, articles and downloaded records, is a central argument in the negotiation on licensing (Bevan et al., 2005). COUNTER statistics enable library managers to empirically assess and shape investment decisions. Without proof of value, the library's profile will weaken.

LIS professionals have to deal with the phenomenon of long-tailed statistics of digital libraries: some intensely used items, and a lot of stuff rarely or never used. And publishers have to explain why and how they sell content on the long tail.

Based on these statistics, new business models emerge that propose for instance a combination of subscription to core collections with a pay-per-view offer for the other items, or even open access to a part of them.

On the other hand, usage statistics provide an in-depth insight into the information seeking behaviour and routines of end users. The CIBER study on scholarly journal usage developed a methodology – deep log analysis – for the evaluation of session patterns and distinguished between different user groups, especially

between repeat and occasional users (Nicholas et al. 2005). Among the analysed patterns are the type of items viewed (list of issues, table of content, abstract, full text HTML and PDF), the median item view time, the day of week, the subject category, the user's geographical location, the place where the journal viewed was published, the number of items viewed in a session, the referrer link (search engine, library, publisher's platform), access through authentication (Athens), as well as attempts to purchase individual items online (pay-per-view).

The significance of these results is that they show what the end users really do. Together with a qualitative survey for the reason why they do so would enable publishers "to deliver more closely to the needs of the user/researcher, hence creating more traffic and more readership, and greater exposure for authors and brands" (Nicholas et al. 2005, p. 278).

But usage statistics provide more information. Like citations, lending and document supply (Salaün et al., 2000), usage statistics may be interpreted as a marker of scientific value of the accessed content. The underlying idea is that "what is used has value".

Unfortunately, little empirical evidence has been published so far on the usage of grey literature in open archives. In the early period of open access initiatives, technical and political aspects prevailed. It was also more important to find sustainable and interoperable solutions than to reflect on the real usefulness, e.g. return on investment.

We reported elsewhere on first results from different repositories (Schöpfel et al., 2009). The figures are consistent: the average download rate of grey items comes out to be higher than for journal articles and other published work. This would highlight the specific value of grey items and their valorisation through open repositories (see also Harnad et al., 2009).

Nevertheless, we should be careful with interpretation. Repository usage statistics are biased by search strategies, accessed content and referring tools. Traffic and readership are enhanced through web citations, and even if we didn't find empirical evidence in published studies thus far, usage statistics are probably linked to web based citations in the way in which the more an item is cited, the greater is the probability that it is used. Also we should keep in mind that compared to academic journals, we know much less about citation patterns and the impact of theses, reports or working papers.

We already mentioned another problem – the poor quality of metadata and the lack of standards for usage statistics and grey literature in repositories. Actually, some projects in the UK, Germany, France and Japan tackle these problems. On the agenda: usage assessment on the item-level, a common terminology, a set of recommendations for repository usage statistics (code of practice), including suggestions for added value services.

15.5 Cost analysis

A short glimpse on the literature confirms Linn's (2009) statement that "it is unfortunate that there are so few good examples of how librarians can use cost-benefit analysis". Estimates of ROI call for budget figures. By capturing cost information for an institutional repository, it would be possible to determine the development cost for one item (full text deposit and/or metadata); over time, it would be possible to link these figures to usage data. But what has become a routine for other kinds of digital libraries (Byrd et al., 2001; Boukacem-Zeghmouri & Schöpfel, 2008) is still largely absent for institutional repositories.

"The costs of digital preservation in general are still difficult to calculate, and it is unclear as yet how much of the work will be funded. It is equally unclear how open-access in general will be funded. Establishing costing and funding models for digital preservation of open-access materials is therefore doubly difficult." (Pinfield & James, 2003).

There is consensus however on one point: "Open Access needs funding" (Lafon, 2010), and "someone has to pay the costs for (…) repositories" (Kennan & Wilson, 2006). No doubt: the institution that produces and hosts a repository has to bear the costs itself.[2] "Institutions have the resources and infrastructure to set up, support and fund repositories" (idem). But what are the cost elements related to repositories? A literature survey[3] uncovers some main cost centers:

Table 3: Cost elements of an open repository

Initial costs	Hardware	**Purchase of server**
	Software	*Uploading*
		Configuration
	Staff	*Project management*
Operating costs	System	*Maintenance*
	Staff	*Metadata production*
		Item selection/validation
		Publicising/promotion
		Attendance at forums etc.
		Negotiating IP rights
	Facilities	*Power*
		Equipment
		Staff floor space

2 This may not be a sustainable business model for all repositories. In January 2010, the Cornell University Library announced a new voluntary, collaborative subscription-like business model to engage institutions that benefit most from arXiv; these institutions should support arXiv through annual contributions to the operating costs. http://arxiv.org/new/#jan2010

3 Granger et al. (2000), Horwood et al. (2004), McDonald (2005), Kennan & Wilson (2006), Piorun & Palmer (2008)

Repository software such as EPrints or DSpace are open source, designed for easy implementation - one day of work for someone experienced with setting up Web servers - so that the major initial cost probably will be the purchase of hardware (Horwood et al., 2004).

McDonald (2005) assessed the amount of $30,500 as startup first year costs for an institutional repository, with more than 60% for staff.

This is compliant with data from the University of London Computing Centre for another project on digital preservation where the staff accounted for 70% of total costs and the next greatest cost was maintenance for hardware and software associated with access (Granger et al., 2000).

Perhaps annual depreciation expense should also be taken into account, during 5-10 years or more, because of the heritage nature of institutional repositories.[4]

Depending on the project, other tasks may include identifying metadata elements, obtaining and tracking permissions, scanning of documents and workflow coordination. Piorun & Palmer (2008) reported on the creation of an institutional repository with initially 320 theses. They estimated the processing costs for each item (digitizing, uploading) at around $70, with an average processing time of 170 minutes per item.

Willinsky (2006) stated that the annual funding of the best known e-print archive, arXiv.org, was $300,000 prior to its move to Cornell University in 2001, corresponding to costs of $9 per paper. The arXiv currently costs $400,000/year, with costs projected to reach $500,000 in 2012[5], corresponding to an annual increase of 5-10% and an average cost per item of about $7[6]. The French HAL archive was told to bear an annual budget of approx. €200,000. This would correspond to costs of €5 per item for the hosting structure.

It is generally admitted that publishing via an institutional repository is not very expensive, even if the deposit costs are added. With an average deposit time of 15 minutes per item this corresponds to costs of roughly €15 per deposit and metadata creation. Costs are low because of missing peer review procedures.

Nevertheless, even if some figures have been published, information about institutional repository costs is incomplete and a general framework for a cost analysis is (still) missing. In particular, it seems quite difficult to estimate costs in a distributed network of repositories.

15.6 Metrics

Derived from usage statistics, cost analysis and other data, at least six measures can be calculated that provide elements for the assessment of impact and return on

4 Acknowledgement to Gilbert Puech, director of the PERSEE journal archive.

5 See http://scholarlykitchen.sspnet.org/2010/01/21/arxiv-grows-up/

6 See http://openaccess.eprints.org/index.php?/archives/702-Annual-Costs-Per-Deposit-of-Hosting-Refereed-Research-Output-Centrally-Versus-Institutionally.html

investment for institutional repositories (IR), especially in comparison with other repositories and digital libraries.

(1) IR costs per item: What is the part of annual expenditures related to one item? This corresponds to the "cost per article" metrics for serials. Examples following Willinsky (2006) and recent data as shown above:

Table 4: Open repository costs per item (examples)

arXiv 2006	$9
arXiv 2009	$6
arXiv 2012	$8
HAL 2008	€5

The figures for open repositories seem higher than for (commercial) e-journals collections, probably because of the relatively low number of annual deposits in IR. But this indicator evolves over time, and with increasing input and controlled budget this cost indicator would decrease.

(2) Cost per item request: What is the part of annual expenditures related to one item request (in terms of access and download)? This corresponds to the "full-text article requests" metrics for serials. Example: in the case study published by Piorun & Palmer (2008) on an IR of digitized dissertations, the average cost per item request for the first year was around $1,90. This corresponds approximately to usage metrics for e-journals (see Boukacem-Zeghmouri & Schöpfel, 2008). Improved referencing and promotion but also the effect of a critical mass ("long-tail effect") will boost this measure.

(3) Item requests per collection: What is the average access and download number per item in a given collection? This corresponds to the "full text article requests per title" metrics that can be calculated for the whole IR as well as for sub-collections or document types. Some examples for collections of document types:

Table 5: Item requests per collection (examples)
(*Malotaux, 2009; ** Merceur, 2007).

Articles*	40
Articles**	8
Theses*	100
Theses**	70
Reports**	30

The interest in this indicator is that it allows for comparison of usage of different document types (here grey literature vs. published articles), laboratories etc., depending on the particular structure and metadata of an IR. It provides elements for the assessment of interest and usage of specific sections of the IR.

(4) IR costs per user: What is the part of annual expenditures related to an individual user? This links the overall expenditures to the number of users. This measure requires an analysis of the log files and would provide an additional element to the assessment of impact, popularity, and readership. There is no valid data for the IR with cost information.

(5) IR costs per depositing author: What is the part of annual expenditures related to an individual depositing author? This links the overall expenditures to the number of users in terms of depositing authors; and this requires an analysis of metadata and would provide information about the acceptance and use in the institution. No valid data is available.

(6) IR items per scientific output: What is the part of the institution's publications that has been deposited in the IR? This provides an estimation of the part of a given institution's scientific production available through its own institutional repository. Two examples:

Table 6: IR items per scientific output (2003-2007)
(*source: SCImago Institutions Rankings 2009 World Report)

Institution	Output*	IR	%
INSERM (France)	34,235	3,115	9
ETH Zurich	8,886	4,013	45

High rates were reported from institutions with a mandatory policy, like the University of Southampton or ETH Zurich. Yet, accurate data on scientific production, especially of grey literature, are difficult to obtain, especially because of missing metadata. Also, mandatory policies may result in uploading metadata without full text.

The problem with all theses measures is that their value depends largely on the availability and quality of metadata, usage statistics, and cost elements. Actually, it appears much more difficult to obtain precise data on this part of the STI market than for (commercial) digital libraries. If we want to know more on the function and place of non-for profit (grey) literature in this new landscape, these data are badly needed.

Another point is that the cost-related metrics change with the development of IR and the depreciation, e.g. the reduction in the value of the initial investment in hard- and software. Even if these measures are defined for a given period (one year) they could also be calculated in a cumulative way.

Alternatives to this ROI assessment are impact measures derived from ranking (webometrics[7]) or link analysis. But these measures remain on the repository level and do not allow for deeper analysis of IR content, such as grey literature.

7 See http://repositories.webometrics.info/

15.7 Concluding remarks

The question of ROI in institutional repositories renders grey literature more discernible in the global economic reasoning of scientific information. Concretely, associating the concept of ROI and institutional archives could lead to a new business model with grey literature gaining new legitimization.

The actual political framework of research – project funding – is related to the evaluation of institution, and incidentally, of institutional repositories which could become, through a mandatory policy of green or gold road, a kind of grey backup reservoir, an alternative to the big deal business model, which seems to be approaching its limits.

One benefit of an evaluation approach covering institutional repositories would be to strengthen the academic library's integration into the scientific project of the university and to replace the scientific information into the centre of the scientific policy. In this context, institutional archives and grey literature could become a central part of scientific evaluation.

Björnshauge (2006) said that research funders demand quality, figures, and metrics. IR projects have to account for costs and need to guarantee efficiency, accountability, and sustainability. Thus, not only visibility but also impact.

The ongoing PEER project[8], launched by STM publishers and co-funded by the European Union, may provide more evidence on economic impact and financial issues of open archives but the PEER research is limited to mostly English-language journals and doesn't take into account other, unconventional material.

John Houghton (2009) compared costs and potential benefits of open access models for scholarly publishing in the UK, Netherlands, and Denmark. Again, the analysis is limited to academic journals.

Grey literature is not a specific category of document but a specific (non commercial) way of access and dissemination of information. The definition of grey literature is an economic definition, nothing else. With the changing research environment and new channels of scientific communication, it becomes clear that grey literature needs a new conceptual framework. The ROI approach with its cost-benefit-analytical tools contributes to this new theory of grey literature.

Bibliography

S. Bevan, et al. (2005). `How usage statistics can inform national negotiations and strategies'. *Serials: The Journal for the Serials Community* **18**(2):116-123.

L. Björnshauge (2006). `The Next Generation of IRs – enabling closer cooperation & networking'. In *International Workshop on institutional repositories and enhanced and alternative metrics of publication impact. Berlin, Humboldt University. 20-21 February 2006*.

8 Publishing and the Ecology of European Research http://www.peerproject.eu/

C. Boukacem-Zeghmouri & J. Schöpfel (2006). `Document supply and open access: an international survey on grey literature'. *Interlending & Document Supply* **34**(3):96-104.

C. Boukacem-Zeghmouri & J. Schöpfel (2008). `On the usage of e-journals in French universities'. *Serials* **21**(2):121-126.

S. Byrd, et al. (2001). `Cost/benefit analysis for digital library projects: the Virginia Historical Inventory Project (VHI)'. *The Bottom Line: Managing Library Finances* **14**(2):65-75.

S. Chapman (2004). `Counting the Costs of Digital Preservation: Is Repository Storage Affordable?'. *Journal of Digital Information* **4**(2):208-214.

S. Granger, et al. (2000). `Cost Elements of Digital Preservation'.

S. Harnad, et al. (2008). `The Access/Impact Problem and the Green and Gold Roads to Open Access: An Update'. *Serials Review* **34**(1):36-40.

S. Harnad, et al. (2009). `Open Access Repositories - maximizing and measuring research impact through university and research-funder open-access self-archiving mandates'. *Wissenschaftsmanagement* **4**(4):36-41.

M. Heaney (ed.) (2009). *Library statistics for the twenty-first century world : proceedings of the conference held in Montréal on 18-19 August 2008 reporting on the Global Library Statistics Project*. K.G. Saur.

L. Horwood, et al. (2004). `OAI compliant institutional repositories and the role of library staff'. *Library Management* **25**(4/5):170-176.

J. Houghton (2009). `Open Access – What are the economic benefits? A comparison of the United Kingdom, Netherlands and Denmark'. Tech. rep., JISC.

R. Housewright & R. Schonfeld (2008). `Ithakas 2006 Studies of Key Stakeholders in the Digital Transformation in Higher Education'. Tech. rep.

C. Jones (2007). *Institutional Repositories: Content and Culture in an Open Access Environment.* Chandos.

M. A. Kennan & C. Wilson (2006). `Institutional repositories: review and an information systems perspective'. *Library Management* **27**(4/5):236-248.

M. B. Lafon (2010). `Open access to scientific publications'. *Commun. ACM* **53**(2):32-34.

H. Lauridsen (2009). `Are libraries dead?'. In *Ressources électroniques académiques: mesures et usages. Colloque international*.

M. Linn (2009). `Cost-benefit analysis: a disparagement of its misuse and misexplanation'. *The Bottom Line: Managing Library Finances* **22**(3):82-85.

S. Malotaux (2009). `OATAO Archive ouverte multi-établissements. Bilan après un an d'existence'. In *Journées d'études sur les archives ouvertes.* Consortium COUPERIN. 2 et 3 avril 2009.

R. H. McDonald (2003). `D-Scholarship @ FSU'. In *The New Frontier of Institutional Repositories Three Different Libraries, Three Different Plans, One Common Goal.* CNI Fall 2003 Project Briefing Monday December 8, 2003.

R. Melero, et al. (2009). `The situation of open access institutional repositories in Spain: 2009 report'. *Information Research* **14**(4).

F. Merceur (2007). `Gestion d'une archive et d'un moissonneur, l'exemple de l'IFREMER'. In *RPIST 2007,* 20 juin 2007.

S. Pinfield & H. James (2003). `The Digital Preservation of e-Prints'. *D-Lib Magazine* **9**(9).

M. Piorun & L. A. Palmer (2008). `Digitizing Dissertations for an Institutional Repository: A Process and Cost Analysis'. *Journal of the Medical Library Association : JMLA* **96**(3):223-229.

J.-M. Salaün, et al. (2000). `Demand for Scientific Articles and Citations: An Example from the Institut de l'information scientifique et technique (France)'. *Scientometrics* **47**(3):561-588.

J. Schöpfel, et al. (2009). `Usage of grey literature in open archives'. In *GL11 Conference Proceedings. Eleventh International Conference on Grey Literature: The Grey Mosaic: Piecing It All Together. Washington D.C., 14-15 December 2009*.

J. Schöpfel & H. Prost (2009). `Document supply of grey literature and open access: an update'. *Interlending & Document Supply* **37**(4):181-191.

P. Shepherd (2005). `COUNTER 2005: a new code of practice and new applications of COUNTER usage statistics'. *Learned Publishing* **18**(4):287-293.

C. Tenopir (2009). `The Value Gap'. *Library Journal* (15 July 2009).

J. Willinsky (2005). *The Access Principle: The Case for Open Access to Research and Scholarship (Digital Libraries and Electronic Publishing)*. The MIT Press.

Chapter 16

e-Science, Cyberinfrastructure and CRIS

Keith G. Jeffery, Science & Technology Facilities Council, UK
Anne Asserson, University of Bergen, Norway

16.1 Introduction

The last 10 years have witnessed a revolution in the research environment, partly mirrored in the commercial and social environments. The underlying factors concern the increasing price-performance of computer hardware including processing, storage and networks; the improvements in user interface technology including mobile phones making the ICT environment more readily available and - of course – WWW (World Wide Web). Because of the challenges in speeds, volumes and complexity, the research environment tends to anticipate by some years developments in the other environments. The e-Science concept, developed in UK from an initial paper by Keith Jeffery [Je99] encompasses and assumes an e-infrastructure [e-IRG] (in USA cyberinfrastructure [NSFCyb]) consisting of networks, computational servers, data servers and detectors. The e-Science concept, however, builds on this physical layer two more layers; one managing information (derived from data by structuring in context) and surmounted by a knowledge layer recording human-generated knowledge (such as scholarly publications) or computer-generated knowledge (derived through data mining).

Synchronously with e-Science, Anne Asserson, Keith Jeffery and others promoted the concept of CRIS (Current Research Information Systems) and the CERIF (Common European Research Information Format) EU recommendation to member states. CERIF is a rich and flexible data model for CRIS or for interoperation of CRIS with formal syntax and declared semantics – thus making it machine-understandable as well as machine-readable. However, CRIS are a necessary component of e-Science allowing researchers, research managers, educators, entrepreneurs and the media to discover what research is being done, by whom, in which organisations, through which projects, from where the funding comes and what are the outputs including publications, products and patents. Clearly, CRIS form an essential way in the e-research environment - including the e-infrastructure - to index research and make it available. It is common for a CRIS to be associated with a repository of full text (or hypermedia) objects such as scholarly publications i.e. one output of the research. However, the repository

equally may contain grey material such as technical reports which, in fact, may form a large component of the 'know-how' and IP (intellectual property) of an organisation.

This paper argues that the future of grey literature (in the widest sense) lies within the context of an e-research environment populated with CERIF-CRIS and associated repositories.

16.2 The e-Research Environment

In 1998-1999 the UK Research Council community was proposing future programmes for R&D. The author was asked to propose an integrating IT architecture [Je99a]. The proposal was based on concepts including distributed computing, metacomputing, metadata, agent- and broker-based middleware, client-server migrating to three-layer and then peer-to-peer architectures and integrated knowledge-based assists. The novelty lay in the integration of various techniques into one architectural framework [Je04].

The UK Research Council community of researchers was facing several IT-based problems. Their ambitions for scientific discovery included post-genomic discoveries, climate change understanding, oceanographic studies, environmental pollution monitoring and modelling, precise materials science, studies of combustion processes, advanced engineering, pharmaceutical design, and particle physics data handling and simulation. They needed more processor power, more data storage capacity, better analysis and visualisation – all supported by easy-to-use tools controlled through an intuitive user interface.

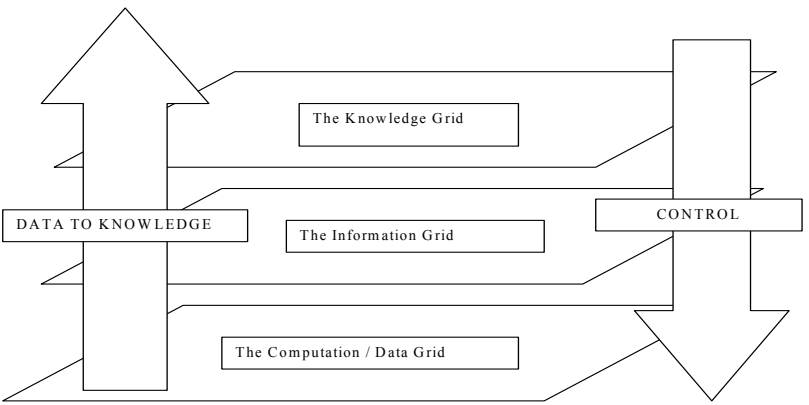

Figure 1: GRIDs architecture

On the other hand, much of commercial ICT (Information and Communication Technology) including process plant control, management information and decision support systems, IT-assisted business processes and their re-engineering, entertainment and media systems and diagnosis support systems all require ever-increasing computational power and expedited information access, ideally through a uniform system providing a seamless information and computation landscape to the end-user. Thus there is a large potential market for GRIDs systems to provide the e-Science (or more broadly e-Research) environment.

The original proposal based the academic development of the GRIDs architecture and facilities on scientific challenging applications, then involving IT companies as the middleware stabilised to produce products which in turn could be taken up by the commercial world. During 2000 the UK e-Science programme was elaborated with funding starting in April 2001.

The architecture proposed consists of three layers (Figure 1). The computation / data grid has supercomputers, large servers, massive data storage facilities and specialised devices and facilities (e.g. for VR (Virtual Reality)) all linked by high-speed networking and forms the lowest layer. The main functions include compute load sharing / algorithm partitioning, resolution of data source addresses, security, replication and message rerouting. This layer also provides connectivity to detectors and instruments. The information grid is superimposed on the computation / data grid and resolves homogeneous access to heterogeneous information sources mainly through the use of metadata and middleware. Finally, the uppermost layer is the knowledge grid that utilises knowledge discovery in database technology to generate knowledge and also allows for representation of knowledge through peer-reviewed scholarly works (publications) and grey literature, especially hyper-linked to information and data to sustain the assertions in the knowledge.

The concept is based on the idea of a uniform landscape within the GRIDs domain, the complexity of which is masked by easy-to-use interfaces. The achievement of this virtualisation is based on metadata [Je00] used in this context [Je04].

16.3 CRIS

CRIS have existed for many decades in research funding organisations and in some research performing institutions. However, it was not until 1991 that experience was shared internationally, although there had been initiatives to interoperate a limited number of CRIS as early as 1984. Driven by various pressure groups, the EC (European Commission) drew together a group of national experts in 1987-1989 to produce the first CERIF (Common European Research Information Format) recommendation. The expert group was reconvened in 1997 to produce the mujch-improved CERIF2000 recommendation upon which all subsequent development is based. In 2002 the EC requested euroCRIS (www.eurocris.org) to take responsibility fore the promotion, maintenance and development of CERIF.

Full details of CERIF are available at www.eurocris.org/cerif. The original purpose of CERIF was to provide a data model for anyone developing a new CRIS and to provide a data model for interoperation between pre-existing (legacy) CRIS (Figure 2). CERIF was developed as a generic datamodel using advanced concepts [AsJeLo2002]. However, CERIF has also been used as a central directory system for an organisation and can be extended further to integrate legacy systems within an organisation [JeAs2006]. CERIF is now becoming more widely used in organisations engaged in R&D whether funders, policymakers, innovators/entrepreneurs, media or academic (research-performing) institutions.

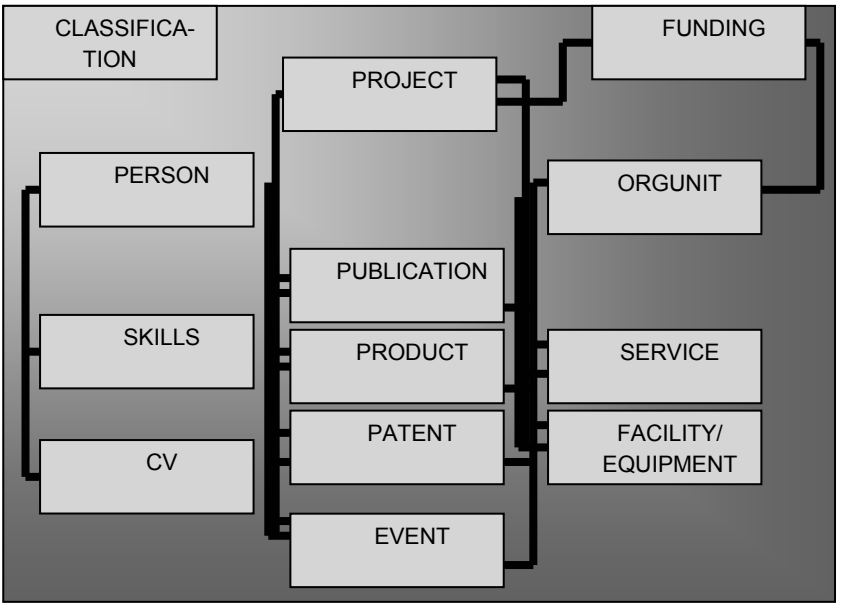

Figure 2: CERIF datamodel

CERIF has some features worth highlighting because of their relevance not only in the research domain but much more widely.

First, CERIF assumes not a hierarchic model of the world but a fully connected (possibly cyclic) graph. This provides great fidelity in representation. For example, many systems have a hierarchic relationship between university department and academic staff member. CERIF can represent accurately an academic staff member related to multiple departments, multiple research groups, multiple academic institutions and commercial organisations.

Second, CERIF separates base relations - as fundamental entities of interest - from relationships. Thus, CERIF has the concept of person (as opposed to researcher, author, employee…) and the role of that person is defined in the relationship of that person to another entity such as an organisation (person P is em-

ployee of Organisation O) or to a publication (person P is author of publication X) or to another person (person P is co-author with person Q) (Figure 3).

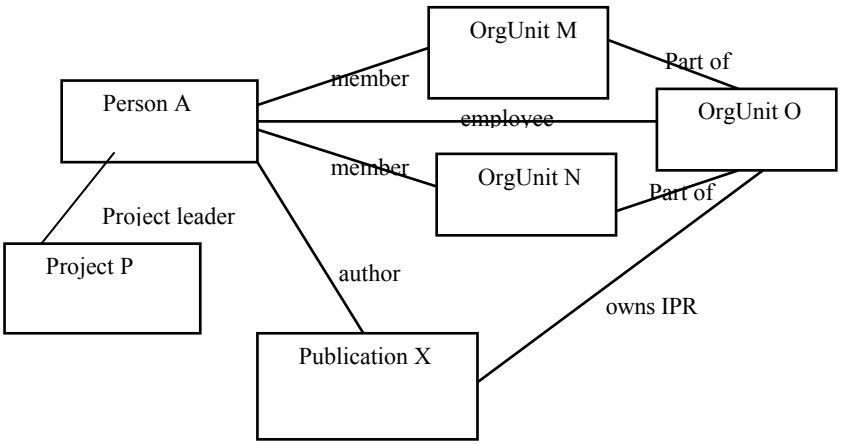

Figure 3: CERIF Relationships

Third, CERIF provides a 'time machine'. This is done not by recording the valid time and transaction time on the base entity instances (the conventional temporal database approach) but by recording date-time-start and date-time-end on the relationship between instances of entities (person P start 20000801:09:00:00 end 20081231:17:00:00 is employee of organisation O). This means it is possible to re-create the history of an instance of an entity (e.g. the CV of a person) or to recreate the state of an organisation (persons, organisational structure, funding, outputs…) at a given date-time or period of time between start datetime and end datetime.

Fourth, CERIF is defined to use Unicode - so that any character set can be represented - and allows declaration of one or more languages for any textual attribute value. This means that multlinguality is handled effectively.

Fifth, CERIF – because of its first-order-logic structure, allows deduction and induction to generate new facts. This is important in saving unnecessary end-user input and permits the support of knowledge-assisted user input and validation.

16.4 Repositories

Repositories (of scholarly material) have developed within a library environment and mainly to record the output of research at an institution – the institutional repository – containing author-deposited copies of peer-reviewed published material; so-called green open access material. This contrasts with gold open access where the author institution pays for a publisher to make the material available under open access on the publisher repository system.

Repositories store and provide access to the detailed information. It is usual – and best practice - to separate repositories of research publications from repositories of research datasets and software (e-Science or, better, e-Research repositories) because of their different access patterns and different metadata requirements. The e-Research repositories require much more detailed metadata to control utilisation of the software and datasets in addition to metadata to allow discovery of the resources. At present they tend to be specific to an individual organisation because of their novelty and the differing requirements on metadata imposed by different (commonly international) communities e.g. in space science, atmospheric physics, materials science, particle physics, humanities or social science.

Publication repositories need not be restricted to peer-reviewed published material. Increasingly institutional repositories include e-preprints and technical reports i.e. grey literature. Some, indeed, include more informal material and teaching material, presentations and lecture notes. Publication repositories typically use some form of Dublin Core Metadata [DC] and most are [OAI-PMH] (Open Archive Initiative – Protocol for Metadata Harvesting) compliant for interoperation and are indexed by Google Scholar. Example software systems are [ePrints], [DSpace], [Fedora] and [ePubs]. Although the metadata associated with the publication includes author name, different publishers / journals / conference proceedings require the name to be in different formats so correlation – and disambiguation from other authors with similar names - is very difficult.

The publication or its metadata may contain information on the institution of the author, but usually only one such organisation even if the author is associated with multiple organisations. Information on the project from which the publication was generated, funding source, facilities or equipment used etc. may or may not be recorded within the publication but not in a structured form and so is more-or-less impossible to extract automatically. Publication repositories require the author to input metadata to describe the publication; this is a threshold barrier and can be reduced by utilising pre-recorded information in the CRIS. The combination of (a) the difficulty of extracting contextual metadata on research as described above from repositories and (b) the threshold barrier caused by human input of metadata leads inevitably to the conclusion that we should link together CRIS and repositories to gain the advantages of each.

Thus, there is an advantage in linking together repositories (with the full text or hypermedia publication and/or repositories with research data and software) with a CRIS which provides structured information on the context of the research – project, equipment, funding, organisations and persons involved [AsJe04]. The metadata in the CRIS describing scholarly publications may be used for evaluation of research; a well-known example is the Norwegian FRIDA [FRIDA] system.

16.5 Organisational ICT

Research funding organisations and research performing institutions need to manage the research. At present most institutions have a complex mix of legacy systems covering this requirement. Worse, commonly they have multiple protocols for intercommunicating with other institutions; an example is the submission of a research proposal from a university to a funding organisation and subsequent transactions involving research products and funding. A CERIF-CRIS can be used as the unifying system [JeAs2006a] over these legacy systems allowing an institution to continue to utilise legacy systems and to replace them as and when business conditions permit. A strong advantage of such a unifying CRIS is that it can be used to support both the workflow of organisational administrative processes and the entry of metadata. The latter involves 'pre-completing' web forms using information stored in the CRIS such as person name, organisation, contact information. Taking the case of a publication, commonly it starts life as grey literature and can be recorded in the CRIS (metadata) and in the repository (full text or hypermedia); if/when it becomes white literature the only additional metadata required concerns the bibliographic information of the publication channel – the remaining metadata information is already stored in the CRIS and thus can be reused [JeAs06b].

16.6 Interoperation

In addition to unifying the IT support of one organisation, the CERIF-CRIS can also be used to interoperate with other institutions thus supporting the distributed and international scale of research – or any commercial / industrial business or social activity. However, interoperation requires a common data format to reduce the many (n*(n-1)) interconversions (between every pair of nodes) to n (each node converts only once to the common standard). There are several architectures to achieve this which were described, characterised and compared [Je05], [JeAs08]: Remote Wrapper; Local Wrapper; Catalog; Catalog plus Pull; Full CERIF; Harvesting. Each has advantages and disadvantages although – obviously – the greatest benefits are obtained by interoperating fully-compliant CERIF-CRIS.

Nonetheless, organisations with legacy systems that are not CERIF-CRIS can utilise one of the techniques mentioned above to 'wrap' their existing system(s) so that interoperation / intercommunication with other organisations utilises CERIF as the canonical information exchange format. Indeed, a special group set up by ESF (European Scince Foundation) at the request of euroHORCS (European Heads of research Councils) reached the same conclusions although euroHORCs decided it was too early for such interoperating systems but encouraged members to converge towards the architecture proposed.

The advantage of interoperating CRIS is that a researcher, research manager, innovator or media reporter can query in a homogeneous way across heterogene-

ous distributed research information sources – including onward access to repositories including peer-reviewed publications and grey literature. It makes possible answers to queries such as 'which researchers are working on drugs to combat HIV/AIDS – sort by country and within country by institution' or 'how many peer-reviewed publications were produced between 1995 and 2000 on global warming – sort by country and within country by aggregated publication impact factor'. In each case the additional access to the repositories via the CRIS used as metadata can provide the full text or hypermedia publication

The benefits are obvious. Researchers can find teams also working in their field – and this is especially important in emerging multidisciplinary fields where the existing subject- or specialism-based academic networks do not yet extend. Research managers can decide on strategy to compete or cooperate with other institutions or – at national scale – with other countries. Innovators can find research ideas relevant to their commercial interests. The media can find 'science stories' that popularise research with the general public and which can stimulate debate on pressing issues – including ethical and funding priority issues – in research; examples include discussions on global warming, GM (genetically modified) food, defence-related research etc.

16.7 Grey in Context

Grey literature – in the widest sense including hypermedia – is produced in the research process and provides a valuable resource. Indeed, commonly it forms the IP of an organisation (technological ideas described) and leads to innovation and wealth creation [JeAs04], or to improved effectiveness and efficiency ('how to' manuals) in the operations of the organisation.

Currently grey literature is usually stored in repositories. There is no common agreement on the metadata to be used to describe this resource although some organisations are using a version of DC and SIGLE [SIGLE} has a defined metadata standard. As suggested [Je99] grey literature (and also white literature) requires richer metadata - than that provided by DC - that has both formal syntax (for efficient computer processing) and declared semantics (to automate processes that would otherwise be performed by humans thus increasing effectiveness and efficiency).

The conclusion was that the metadata should be CERIF-compatible and stored in the CRIS, with the grey literature object – full text or hypermedia – stored in a repository with the two sources linked to allow optimal use of the characteristics of the CRIS and the repository [JeAs05]. In this way not only is the grey literature object provided with better metadata for retrieval but also is associated with the other contextual metadata in the CRIS covering projects, persons, organisations, facilities, equipment, events, products and patents. This truly puts 'grey in context'.

With the CRIS forming the research context backbone information in the e-infrastructure supporting GRIDs and e-research, this further places grey firmly in the research environment together with other publications and products. This architectural approach positions optimally grey literature.

References

[AsJeLo02] Asserson,A; Jeffery,K.G; Lopatenko,A: 'CERIF: Past, Present and Future' in Adamczak,W & Nase,A [Eds]: Proceedings CRIS2002 6th International Conference on Current Research Information Systems; Kassel University Press ISBN 3-0331146-844 pp 33-40

[AsJe04] Asserson, A; Jeffery, K.G.; 'Research Output Publications and CRIS' in A Nase, G van Grootel [Eds] Proceedings CRIS2004 Conference, Leuven University Press ISBN 90 5867 3839 May 2004 pp 29-40 [available under www.eurocris.org]

[CERIF] Commpn European Research Informaiton Format page within euroCRIS website http://www.eurocris.org/cerif

[DC] Dublin Core Metadata Stancdard homepage http://www.oclc.org:5046/research/dublin_core/

[DSpace] DSpace repository product homepage http://www.dspace.org/

[e-IRG] EU reflection group roadmap: http://www.e-irg.org/roadmap/eIRG-roadmap.pdf

[ePubs] ePubs repository product homepage http://epubs.cclrc.ac.uk/

[eprints] ePrints repository product homepage http://www.eprints.org/

[Fedora] Fedora repository product homepage http://www.fedora.info/

[FRIDA] Norwegian FRIDA CRIS Homepage http://frida.uio.no

[Je99a] Jeffery, K.G 'Knowledge, Information and Data' September 1999 Paper submitted to Director General of Research Councils proposing the programme that became e-Science; available at http://www.semanticgrid.org/docs/KnowledgeInformatio-Data/KnowledgeInformationData.html

[Je99b] [Jeffery 1999] Jeffery, K G: 'An Architecture for Grey Literature in a R&D Context' Proceedings GL'99 [Grey Literature] Conference Washington DC October 1999 http://www.konbib.nl/greynet/frame4.htm

[Je00] [Jeffery, 2000]. Jeffery, K G: 'Metadata': in Brinkkemper,J; Lindencrona,E; Solvberg,A [Eds]: 'Information Systems Engineering' Springer Verlag, London 2000. ISBN 1-85233-317-0.

[Je04] Jeffery, K.G.; 'GRIDs, Databases and Information Systems Engineering Research' in Bertino,E; Christodoulakis,S; Plexousakis,D; Christophies,V; Koubarakis,M; Bohm,K; Ferrari,E [Eds] Advances in Database Technology - EDBT 2004 Springer LNCS2992 pp3-16 ISBN 3-540-21200-0 March 2004

[Je04a] Jeffery, K.G.; 'The New Technologies: can CRISs Benefit' in A Nase, G van Grootel [Eds] Proceedings CRIS2004 Conference, Leuven University Press ISBN 90 5867 3839 May 2004 pp 77-88 [available under www.eurocris.org]

[Je05] K G Jeffery CRISs, Architectures and CERIF CCLRC-RAL Technical Report RAL-TR-2005-003 [2005]

[JeAs04] K G Jeffery, A G S Asserson; Relating Intellectual Property Products to the Corporate Context; Proceedings Grey Literature 6 Conference, New York, December 2004; TextRelease; ISBN 90-77484-03-5

[JeAs05] K G Jeffery, A G S Asserson 'Grey in the R&D Process'; Proceedings Grey Literature 7 Conference, Nancy, December 2005; TextRelease; ISBN 90-77484-06-X ISSN 1386-2316

[JeAs06a]Keith G Jeffery, Anne Asserson: 'CRIS Central Relating Information System' in Anne Gams Steine Asserson, Eduard J Simons [Eds] 'Enabling Interaction and Quality: Beyond he Hanseatic League'; Proceedings 8th International Conference on Current Research Information Systems CRIS2006 Conference, Bergen, May 2006 pp109-120 Leuven University Press ISBN 978 90 5867 536 1

[JeAs06b] Keith G Jeffery, Anne Asserson: 'Supporting the Research Process with a CRIS' in Anne Gams Steine Asserson, Eduard J Simons [Eds] 'Enabling Interaction and Quality: Beyond the Hanseatic League'; Proceedings 8th International Conference on Current Research Information Systems CRIS2006 Conference, Bergen, May 2006 pp 121-130 Leuven University Press ISBN 978 90 5867 536 1

[JeAs06c] Keith G Jeffery, Anne Asserson: 'Hyperactive Grey Objects' Proceedings Grey Literature 8 Conference, New Orleans, December 2006; TextRelease; ISBN 90-77484-08-6. ISSN 1386-2316 ; No. 8-06-X

[JeAs07] Keith G Jeffery, Anne Asserson: 'Greyscape' Opening Paper in Proceedings Grey Literature 9 Conference Antwerp [GL9] 10-11 December 2007 pp9-14; Textrelease, Amsterdam; ISSN 1386-2316

[JeAs08a] Keith G Jeffery, Anne Asserson: 'CRIS: Research Organisation View of the e-Infrastructure' in Ales Bosnjak, Maximilian Stempfhuber [Eds] Proceedings CRIS2008 Conference, Maribor, Slovenia June 2008 pp149-158 IZUM, Maribor, Slovenia, ISBN 978-961-6133-38-8.

[JeAs08b] Keith G Jeffery, Anne Asserson: 'INTEREST' Proceedings Grey Literature Conference Amsterdam 8-9 December 2008 in Tenth International Conference on Grey Literature : Designing the Grey Grid for Information Society, 8-9 December 2008, Science Park Amsterdam, The Netherlands ed. by Dominic J. Farace and Jerry Frantzen ; GreyNet, Grey Literature Network Service. - Amsterdam : TextRelease, February 2009. GL-Conference series, ISSN 1386-2316; No. 10. - ISBN 978-90-77484-11-1.

[JeLoAs02] Jeffery,K.G; Lopatenko,A; Asserson,A: 'Comparative Study of Metadata for Scientific Information: The Place of CERIF in CRISs and Scientific Repositories' in Adamczak,W & Nase,A [Eds]: Proceedings CRIS2002 6th International Conference on Current Research Information Systems; Kassel University Press ISBN 3-0331146-844 pp 77-86

[LoAsJe02] Lopatenko,A Asserson,A; Jeffery,K.G;: 'CERIF: Information Retrieval of Research Information in a Distributed Heterogeneous Environment' in Adamczak,W & Nase,A [Eds]: Proceedings CRIS2002 6th International Conference on Current Research Information Systems; Kassel University Press ISBN 3-0331146-844 pp 59-68

[NSFCyb] NSF CyberInfrastructire home page http://www.adec.edu/nsf/nsfcyberinfrastructure.html

[OAI-PMH] Open Archives Initiative Protocol for Metadata Harvesting Homepage

[RDF] W3C Resource Description Framework homepage http://www.w3.org/RDF/

[SIGLE]Open SIGLE Homepage http://opensigle.inist.fr/

[XML] W3C eXtended Markup Language Homepage http://www.w3.org/XML/

Chapter 17

Course and Learning Objective in the Teaching of Grey Literature: The Role of Library and Information Science Education

Debbie L. Rabina, Pratt Institute, USA

17.1 Introduction

A study of grey literature in the context of scholarly communication is intrinsically related to the role of grey literature in the knowledge chain and as part of the changing landscape of knowledge dissemination. The established role of libraries as agents of dissemination of scholarly content was, until the advent of digital libraries, guided by commercial vendors who perceived libraries as a venue to promote their business models. Commercial vendors encourage and promote the dependence of libraries on vendors for subscription content (Allardice, 1997). In recent years, increase in subscription costs and low usage by library patrons (University of California), have librarians looking elsewhere for high quality content, paving the way for grey literature to play a more prominent role in collection development.

While the role of libraries in the dissemination of scholarly content has been addressed in the literature (Mackenzie Owen 2002) the responsibility of library and information science (LIS) schools in the knowledge chain with regard to grey literature, has not received much attention. In order to better understand how LIS schools are preparing future information professionals to work with grey literature, a preliminary survey was conducted in 2007 (Rabina, 2008). This research updates results from the 2007 study and further situates grey literature in the landscape of scholarly communication. More specifically, this study asks the following research questions:

RQ1: Is education for grey literature in LIS education in North America non-specific and embedded within larger topic themes or does it receive unique treatment in the curriculum?

RQ2: Are LIS students aware of grey literature and can accurately describe it?

RQ3: Within the grey literature community, who are the most likely disseminators of education for grey literature among LIS students?

17.2 Development and maturation of grey literature as a scientific discipline

The most-cited definition for grey literature is "that which is produced by government, academics, business, and industries, both in print and electronic formats, but which is not controlled by commercial publishing interests and where publishing is not the primary activity of the organization" (Farace, 1998). ODLIS (Online Dictionary of Library and Information Science) provides a slightly broader definition focusing on the essence of the literature rather than on its origin: "Documentary material in print and electronic formats, such as reports, preprints, internal documents (memoranda, newsletters, market surveys, etc.), theses and dissertations, conference proceedings, technical specification and standards, trade literature etc., not readily available through regular market channels because it was never commercially published/listed or was not widely distributed" (Reitz, 2004). While definitions proliferate, there is agreement on the main characteristics of grey literature: they are materials that are published by entities whose core interests are not in publishing and, as a result, are typically not marketed or distributed by commercial publishing organizations (Mackenzie Owen, 1997). In summary, grey literature is discussed in terms of its origins, its methods of dissemination, or both.

The research conducted by Sulouff et al. (2005), whose paper is most closely related to the theme of this study, points out that grey literature "takes different forms in different departmental settings" so that a working definition is often based on circumstance. The library sector carries responsibility for the management and processing of grey literature. This role is acknowledged by several researchers (Mackenzie Owen, 1997; Sulouff et al, 2005) although they have written largely about the role that librarians take with regard to grey literature, but little about how librarians learn about grey literature. The role of librarians is described as promoting dissemination and use of grey literature through cataloging, searching, archiving and preservation (Mackenzie Owen, 1997). Gelfand believes that these roles, at least with regard to grey literature, are learned on the job: "training and bibliographic familiarity… does not follow a curriculum or a set of readers or textbooks, but instead studies by doing (Gelfand, 1998).

Research regarding grey literature in libraries has focused more on case studies in particular libraries (see Aina, 2000) than on grey literature in LIS education. A review of LIS syllabi, described in more detail below, supports Gelfand's view that education in grey literature is mostly field, and not curriculum, driven.

Thomas Kuhn's theory of the structure of scientific revolutions argues that the point when scientific disciplines change and a paradigm shift occurs within them is the point in which the existing paradigm can no longer account for the observed phenomena taking place with it (Kuhn, 1996). Kuhn's framework has been applied in the library and information science field to identify paradigm shifts in the research and teaching of LIS areas (Richardson, 1986; Smiraglia and Leazer, 1994). Kuhn signals the textbook as a tool that has served as a staple since the

nineteenth century of establishment of a scientific field noting that textbooks "expound the body of accepted theory, illustrate many or all of its successful applications, and compare these applications with exemplary observations and experiments (Kuhn, 1996, p. 10). As a relatively young field of research, grey literature has not developed an established textbook and instructional curriculum, but this should not be interpreted as lack of establishment of the field, but rather as an indication of the field's adaptability, particularly in an era where textbooks are being criticized and their sales declining (Howard, 2008).

Library and information professionals are a vital link in the chain that makes grey literature available to researchers, students and the interested public. While on-the-job training is invaluable, the purpose of graduate-level training is that professionals are hired with some baseline knowledge that they bring to the workplace upon graduation. Courses that educate future information professionals in areas relating to grey literature are critical training ground if awareness to grey literature is to increase.

This study aims to identify what students currently enrolled in LIS graduate programs know about grey literature and where they are learning it. Once we have a clearer picture of the training currently available, we can open a discussion between LIS professionals, LIS educators and LIS students to determine how LIS education can best assist in meeting the needs of the current workplace and use LIS education to strengthen the relevance of current graduates to the workplace.

17.3 Methodology and data collection

To gauge the current place of grey literature in library and information science education data was collected by several means and from several sources. The first research question, asking whether education for grey literature in LIS education in North America non-specific and embedded within larger topic themes or does it receive unique treatment in the curriculum, was tested by means of course review from the top ten LIS programs in the United States.

The second research question, asking about LIS students' awareness of and knowledge about grey literature, was tested using a closed-form questionnaire, and finally, the third research question, asking who are the most likely disseminators of education for grey literature among LIS students, was tested by examining the bibliometric output of presenters in the grey literature conferences.

To understand where grey literature fits within the courses offered at LIS program, the researcher examined course descriptions and syllabi of the 2009 top ten LIS graduate programs in the United States (U.S. News and World Report, 2009). Data collection from syllabi is often limited by publication practices and policies of individual LIS programs. There is a very wide range of materials available from different programs, from those programs and/or professors that make all syllabi, slides, and notes available on the course open website, to those that provide only a short course description and make syllabi available only through password pro-

tected learning managements systems (such as Moodle, Blackboard, etc.). Data for this research was collected from all sources available at each of the LIS programs reviewed, which included in all cases course descriptions from the university's official bulletin, and in some cases, syllabi for individual courses. In addition, the research interests and publications of faculty members in each school were reviewed to identify faculty with research interests in grey literature.

Students' awareness of and knowledge about grey literature was assessed by administrating a closed-form questionnaire to LIS students at a mid-size urban LIS program in the United States. The questionnaire was distributed in hard copy during June 2009 in classrooms. In total forty-eight questionnaires were collected with a response rate of 100%. The survey contained four questions in which students were asked of their knowledge about grey literature and where this knowledge was obtained. Data from the completed questionnaires was entered into an online survey program for further analysis. Limitations of surveys as a data collection method are inherent in the instrument; results are self reported and could be skewed by intentional deception, misinterpretation of the questions, and a desire to please the researcher. To avoid these limitations to the greatest degree possible, the questionnaire was tested for reliability in a pilot study conducted with a small group of students during late May 2009 and the final version was based on their feedback.

To identify the main agents of dissemination of research and scholarship about grey literature, bibliometric data was collected about the output of researchers publishing in the area of grey literature. Data was collected for one hundred and three researchers who have published in the first four volumes of *The Grey Journal*. Data included role and affiliation of each researcher (librarian, researcher, LIS faculty member), extent of LIS teaching activity (part time, full time or none), total number of publications in *The Grey Journal*, total number of publications in other journals, and h-index of the researcher. Researchers were awarded points for each of these factors with the highest scores identifying the likeliest disseminators of information about grey literature. Data about publications in *The Grey Journal* was collected from the table-of-contents pages. Data about other publications was collected from three journal databases (Library Literature, Emerald and Library and Information Science Abstracts). Data about journal impact factor was collected from Web of Science and finally, the h-index was taken from Scopus.

Limitations of data collection for testing the research question include difficulties in identifying authors by name (e.g., there may be several authors with the same name) and difficulties in establishing the teaching statues of each of the researchers in the study. By using multiple sources and databases the researcher tried to achieve the most accurate results possible. An additional limitation is that data was collected only from traditional forums of scholarly communication such as articles and conference proceedings, and did not address web 2.0 forums such as blogs or professional forums, and it is entirely possible that individuals who communicate about grey literature through blogs or listservs contribute to grey literature education in not insignificant ways.

17.4 Results

In order to test the first research question, asking whether education for grey literature in LIS education in North America is non-specific and embedded within larger topic themes, the researcher examined course syllabi, faculty publications and faculty research interests at the top ten LIS programs in the United States, to see how prominent a presented grey literature has in each one on these indicators. Results showed very little activity in all these areas. No courses devoted to grey literature were identified and no courses specifically mentioned grey literature in the course description, or where available, in course syllabi. Very few faculty members in these schools conduct research in the area of grey literature.

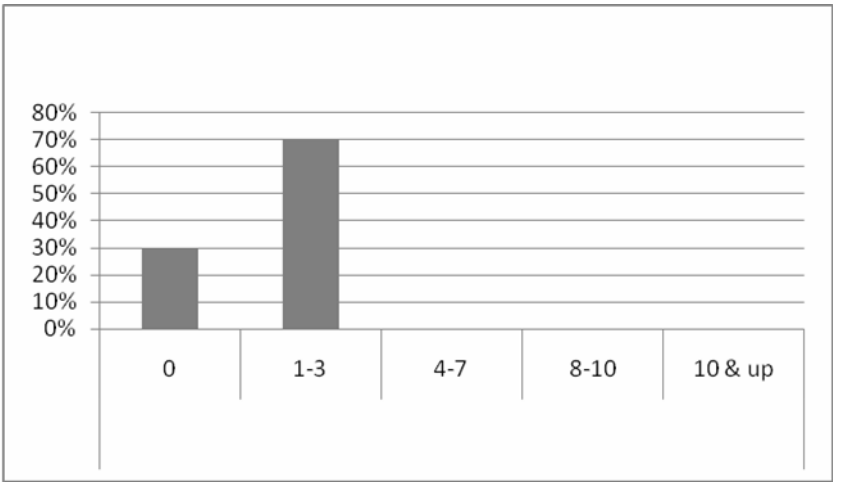

Figure 1: Occurrences of GL in total (including syllabi, research interest, publications, etc.) in top-ten LIS programs in North America

These results indicate that education for grey literature in not specific, i.e., not offered in designated courses. The extent to which education about grey literature in covered in other courses such as collection development, knowledge organization or special collections, could not be fully determined from the information available from the population studied. Only by drawing on the results of students' questionnaire, indicating overall familiarity with the term 'grey literature' can we assume the LIS education covers the concepts and characteristics of grey literature in some courses offered in LIS programs.

The second research question, examining whether despite the lack of structure in education for grey literature, most LIS students are aware of its existence and can accurately describe it was tested by administrating a questionnaire to students, as described in the section above. Of the responses collected, 54.2% of respondents indicated that they had heard of the terms grey literature. These results are significantly higher from the 25% found in earlier research (Rabina, 2008) and are

attributed to the distribution of the questionnaire in a smaller number of LIS schools than in the earlier study. While the result found in the 2007 study is likely more true to reality, the high result of the 2009 study is supported by the findings of the third research question below.

In order to determine if students understand the nature of grey literature, they were ask to read ten statements and indicate how well the statements describe grey literature. Responses were on a Likert scale with 5 meaning the statement described grey literature very well and 1 meaning that it does not describe it well. The results, in table 17.1, indicate that students accurately identify grey literature and recognize its main characteristics.

Table 1: How well does each of the following statements describe grey literature?

Very well	Statement	Not well
58.6%	Grey literature are materials not indexed by commercial indexers	20.6%
50%	Grey literature describes materials published by non-commercial publishers	23.4%
43.4%	Grey literature describes materials not available in OPACs	19.2%
39.9%	Grey literature describes materials of unknown origin (where the author or publisher can't be identified)	46.5%
24.1%	Grey literature refers to any ephemeral materials	62%
20.7%	Grey literature describes materials not picked by commercial search engines (such as Google and Yahoo)	44.8%
16.7%	Grey literature is similar to open access journals	66.6%
14.3%	Grey literature refers to materials guarded by institutional gatekeepers who deny access to them	75%
10.7%	Grey literature is government information that is not available in the Catalog of Government Publications	53.6%
7.1%	Grey literature refers to materials stored in dark archives that are intended for long term preservation	92.9%

The third research question, asking who within the grey literature community are the most active disseminators of education for grey, was answered by evaluating the research productivity (number of publications, citations, and h-index) of individuals publishing in The Grey Journal, as well as their teaching activity. Each author was given points for research productivity and points for teaching activity. The results, in table 17.2 indicate there is a correlation between teaching and research activity: full time teachers are engaged in higher volume research compared to non or part time teachers.

Table 2: Correlation between teaching and research productivity (N=99 ; shaded area=correlation)

	Non teachers	Part-time teachers	Full-time teachers
Low research activity	78.7%	50%	7.1%
Medium research activity	13.1%	29.2%	0%
High research activity	8.2%	20.8%	92.9%

17.5 Discussion

The results of this study clarify the state of LIS education regarding grey literature. Regarding the prevalence of grey literature in master's level programs of library and information science in the United States, results indicate that grey literature receives little attention in the curriculum. In the master's programs examined, no courses dealing with grey literature were identified and very few occurrences of the term within course materials or scholarly activity within the maters' program were found. The current situation implies that thorough knowledge and working practices with grey literature are acquired in the workplace and not through graduate course work.

In spite of scant evidence of teaching grey literature, a large number of students surveyed were able to correctly describe that nature and characteristics of grey literature, indicating that notwithstanding the lack of structure in education for grey literature, most LIS students are aware of its existence and can accurately describe it. Students perceive grey literature as lacking in bibliographic control (not indexed by commercial indexers and not available in online public access catalogs) and created by non-commercial publishers.

This finding suggests that the scope and depth of knowledge acquired throughout the master's program, allows students to make informed judgments regarding the accuracy of the statements provided in the questionnaire.

The third research question, asking who are the strongest disseminators of grey literature education within the grey literature community, indicate those engaged in teaching are likely to be engaged in high research volume, and are most likely to be powerful agents for teaching future information professionals about grey literature. Knowledge is disseminated in academia through scholarly activities that include teaching and research. While many engage in one or the other, those engaged in both are positioned to have the greatest impact. The data confirms a correlation between the two variables – individuals engaged in full time teaching activity are also engaged in high volume research activity. These two venues, teaching and publishing, provide the opportunity to reach a wide audience

These findings can assist LIS educators in increasing students' knowledge of grey literature and help establish best practices for grey literature education.

17.6 Recommendations for best practices for grey literature education

When identifying gaps in LIS education, the more common approach has been to suggest and outline a suitable course curriculum for that topic (Heintz, 2004; Weimer and Reehling, 2006), but there are several arguments to be made in favor of a cross curricular approach for grey literature education, mainly, the opportunity to expose more students to grey literature than would be possible through a designated course. The cross-curricular approach to teaching grey literature in accordance with the interdisciplinary scope of grey literature content.

A cross-curricular approach to grey literature education is best offered in several courses, including some that are traditionally part of schools' core offerings, such as knowledge organization and reference, as well as courses that are usually offered as electives, such as collection development and specialized reference courses (for example, scientific information sources, government information sources, statistical information, health information and more). Distribution across the curriculum will address the main areas of importance to library and information professionals dealing with grey literature on two levels: working with the public and working behind the scene. Working with the public addresses the question of the grey literature needed by reference librarians for their work with library patrons seeking information in all areas, whether health information, scientific information or information in the arts and humanities. Working behind the scenes will address questions about the best ways to locate grey literature, to gain bibliographic control over it, to incorporate it in the library's OPAC, website, subject guides and more.

References

Aina, L.O. (2000). Grey literature and library and information studies: A global perspective. – In: The International Journal of Grey Literature 1 (4), pp. 179-182.

Allardice, C (1997) Subscription agents, libraries and others involved in the information dissemination, Electronic Publishing '97 - New Models and Opportunities: - In: Proceedings of an ICCC/IFIP conference held at the University of Kent, Kenterbury, UK, April 14-16 1997/ Edited by Fytton Rowland and Jack Meadows / ISBN 1-891365-00-2 http://elpub.scix.net/cgi-bin/works/Show?97129 (http://elpub.scix.net/data/works/att/97129.content.pdf)

Farace, D. (1998). Forword - In: GL '97 Conference Proceedings: Third International Conference on Grey Literature: Perspectives on the Design and Transfer of Scientific and

Technical Information, 13-14 November 1997. Amsterdam: Transatlantic/GreyNet. – iii p. (GL conference series; no. 3).

Heintz, J. P. (2004). Developing a library school course in government statistics. – In: Journal of Education for Library and Information Science, 45 (3), pp 189-199.

Howard, J. (19 Sept. 2008). Textbook sales drop, and university presses search for reasons. – In: The Chronicle of Higher Education 55 (4), p. A10.

Gelfand, J. (1998). Teaching and exposing grey literature: What the information profession needs to know – example from sciences. – In: Collection Building 17 (4), pp. 159-166.

Kuhn, T. (1996). The structure of scientific revolutions. University of Chicago Press.

Mackenzie Owen, J. (1997). The expanding horizon of grey literature. - In: Third International Conference on Grey Literature, Luxembourg.

Mackenzie Owen, J. (2002). – In: The new dissemination of knowledge: Digital libraries and institutional roles in scholarly publishing. Journal of Economic Methodology 9 (3), pp. 275-288

Rabina, D. L. (2008) A scholarly communication perspective of grey literature in library and information science education. – In: The Grey Journal 4(1), pp. 34-40.

Reitz, Joan M. (2004). ODLIS – Online Dictionary for Library and Information Science. Libraries Unlimited http://lu.com/odlis/index.cfm

Richardson, J. V. (1986). Paradigmatic shifts in the teaching of government publications, 1895-1985. – In: Journal of Education for Library and Information Science 24 (6), pp. 249-266.

Smiraglia, R.P., G.H. Leazer (1994). Reflecting the maturation of a profession: Thirty-five years of Library Resources and Technical Services. – In: Library Recourses and Technical Services 38, pp. 27-46.

Sulouff, P., S. Bell, J. Briden, S. Fronyz, A. Marshall (2005). Learning about grey literature in interviewing subject librarians: A study at the University of Rochester. – In: *College and Research Libraries News 66* (7) http://www.ala.org/ala/mgrps/divs/acrl/publications/crlnews/2005/jul/learngreylit.cfm

University of California. Reshaping Scholarly Communication (2004) http://osc.universityofcalifornia.edu/journals/journals_a.html

U.S. News and World Report, 2009. Ranking, library and information studies. http://gradschools.usnews.rankingsandreviews.com/best-graduate-schools/top-library-information-science-programs/rankings

Weimer, K.H., P. Reehling (2006). A new model of geographic information librarianship: Description, curriculum and program proposal. – In: Journal of Education for Library and Information Science 47 (4), pp. 291-302.

Appendices

This monograph contains five appendices that may well help in understanding, learning, researching, and accessing grey literature.

Appendix I is a compilation of biographical notes provided by the authors in this monograph. More information on authors in grey literature can be found on the WHOIS webpage of TextRelease, the Program and Conference Bureau for the International Series on Grey Literature.

Appendix II provides examples of grey literature and profiles organizations responsible for its production and/or processing. Only web-based resources that explicitly refer to the term grey literature (or its equivalent in any language) are listed. The web-based resources appear within categories derived from the CO-SATI (American) and SIGLE (European) Classification Systems.

Appendix III produces a list of grey document types that was first compiled in 2004 during a study on citation analysis and grey literature. Since then, this list has been maintained on GreyNet's website and further developed by the international grey literature community. It is interesting in that it illustrates the wide range and heterogeneity of grey literature.

Appendix IV provides the titles of volumes in the International Conference Series on Grey Literature from 1993 to 2010 along with links to these collections available in the OpenSIGLE Repository.

Appendix V provides the thematic titles of the volume/issues in The Grey Journal from 2005 to 2010. The Grey Journal (TGJ) is currently the only international journal on grey literature and is published by TextRelease in Amsterdam. TGJ is indexed in the Scopus database as well as by other A&I services.

Appendix I
Biographical Notes on the Authors

Anne Asserson holds a Cand. polit. with a Masters in Information Science from the University of Bergen, UiB. She has been working with Research Documentation, and has participated in substantial parts of CRIS developmental work, locally and nationally. Anne Asserson has been part of the establishing and implementing of a Research Documentation system, Fdok http://www.ub.uib.no/fdok/sok/, at the UiB. For several years she was the chairwoman of the Steering Group of the national CRIS system and project secretary of a National system for academic administration. Anne Asserson is presently representing UiB in the national group that is implementing a new national research documentation system, FRIDA. She has also participated in The CORDIS funded European-wide project on " Best Practice" 1996 . She was a member of the working group set up 1997 that produced the report CERIF2000 Guidelines (1999) www.cordis.lu/cerif, coordinated by the DGXIII-D4. euroCRIS is now the custodian of the CERIF model www.eurocris.org. Anne Asserson is a member of the Best Practice Task Group. Email: anne.asserson@fa.uib.no

Marcus A. Banks is a librarian at the New York University School of Medicine. He is also the editor of the open access journal Biomedical Digital Libraries, and the Chair of the Medical Library Association's Task Force on Librarians without Borders. Marcus is also the recipient of the GreyNet Award 2006. He is interested in how the concept of grey literature will continue to evolve in the digital age. Email: mab992@yahoo.com

Chérifa Boukacem graduated with a Master's Degree in library sciences from the University of Algiers (DZ) in 1996. She obtained her Ph.D. from the University of Lyon-2 in 2004 with a comparative study on French academic interlibrary loan and access to digital resources. From 2001 to 2003, she worked at the department of scientific and technical information of the University of Lyon-1, she teaches since 2000 library and information science to students and professionals, is member of the French research group in library sciences, DOCSI/GRESI, and co-edits a digital serials, Konex. Her main interests are on e-publishing and multimedia, library software, statistics and Scientometrics and professional counseling. Actually, she is working on the transformation of the digital library and on a longitudinal study of document supply, impact factor and the switch from print to electronic resources. Email: boukacemc@yahoo.fr

Todd A. Chavez is Director of Collection Analysis for the University of South Florida Libraries. His research agenda includes assessment of collections advancing interdisciplinary research and Internet-based subject/discipline knowledge portals. Current projects include the Karst Information Portal and a medieval studies portal connecting Florida's emerging community of medievalists. Email: tchavez@lib.usf.edu

June Crowe is the Senior Researcher and Group Manager, Open Source Research Division at Information International Associates, Inc. (IIa). She received her AMLS from the University of Michigan, Ann Arbor and her M.Ed. in geographic education from the University of Georgia, Athens. She has extensive experience in the management and operations of library services across government, public, academic, and special libraries. At IIa she manages the open source research division which focuses on medical, socio-cultural, science and technology and business research. Her primary interests are open source information in Grey Literature, digital repositories, and open source intelligence tools. Email: jcrowe@iiaweb.com

Dominic J. Farace is Director of TextRelease, an Amsterdam based information bureau specializing in grey literature and networked information. He is a native Louisianan and holds two degrees in sociology from Creighton University (BA) and the University of New Orleans (MA). His doctoral dissertation in social sciences is from the University of Utrecht, The Netherlands, where he has lived and worked for the past twenty-seven years. After six years heading the Department of Documentary Information at the Royal Netherlands Academy of Arts and Sciences (Swidoc/KNAW), he founded GreyNet, Grey Literature Network Service, in 1993 and has since been responsible for the international GL-Conference Series. In this capacity, he serves as Program and Conference Director as well as managing editor of the conference proceedings. Since 2004, he is a Guest Lecturer on Grey Literature in the Masters Program at the University of Amsterdam; Instructor of Grey Literature via UNO Distance Education, and Editor of TGJ, The Grey Journal. Email: dominic.farace@textrelease.com

Jerry Frantzen graduated in 1999 from the College of Amsterdam in Library and Information Science. Frantzen is the technical editor of The Grey Journal (TGJ). And, since 1996, he is affiliated with GreyNet, Grey Literature Network Service, as a freelance technical consultant. Email: info@greynet.org

Anne Gentil-Beccot studied French literature before obtaining a master degree in Information Science. She is now serials librarian in the CERN Scientific Information Service and contributes to the CERN Open Access actions. Email: Anne.Gentil-Beccot@cern.ch

Nathalie Henrot graduated in History, then in Information Sciences from the University of Tours in 1988. She has been working for the INIST-CNRS for seventeen years, more specifically at the Monographs & Grey Literature Section from

1993, for congress proceedings acquisition. She is now the user administrator in the OpenSIGLE project. Email: henrotn@inist.fr

Gail Hodge, IIa Senior Information Scientist, is an authority in the field of information science, with an emphasis on metadata, taxonomy, and thesaurus development. For over 25 years, she has helped organizations develop systems for capturing metadata and for using classification schemes, thesauri, and other knowledge organization sources in a variety of fields, including the environment, life sciences, and aerospace. Email: ghodge@iiaweb.com

Keith Jeffery is currently Director, IT and International Strategy of STFC (Science and Technology Facilities Council), based at Rutherford Appleton Laboratory in UK. Previously he was Head of Business and Information Technology Department with a staff of 140 supporting over 360000 users, developing software for business and science and doing leading edge R&D. STFC hosts the UK and Ireland Office of W3C and develops and supports the largest OA (Open Access) institutional repository in UK. Keith is a Fellow of both the Geological Society of London and the British Computer Society. He is a Chartered Engineer. He is an Honorary Fellow of the Irish Computer Society. He is president of euroCRIS (www.eurocris.org) and of ERCIM (www.ercim.org) and holds three honorary professorships. He has extensive publications and has served on numerous programme committees and research grant review panels. He has particular interests in 'the research process' and the relationship of hypotheses, experiments, primary data and publications based on research in information systems, knowledge-based systems and metadata. Email: k.g.jeffery@rl.ac.uk

Primož Južnič is an associate professor at the Department of Library and Information Science and Book Studies at Faculty of Arts, University of Ljubljana (Slovenia). His main area of research and interest is bibliometrics, collection management and LIS education. He teaches the following courses: Bibliometrics, Special libraries, and Collection Management. Before starting his university career, he was a heading different special and academic libraries and information/computer centres. He was also working at the European Commission, for three years, as the seconded informatics expert. Email primoz.juznic@ff.uni-lj.si

Tomas Lipinski obtained his J.D. from Marquette University Law School, LL.M. from The John Marshall Law School, and Ph.D. from the University of Illinois at Urbana-Champaign. Professor Lipinski has worked in a variety of library and legal settings including the private, public and non-profit sectors. Professor Lipinski teaches researches and speaks frequently on various topics within the areas of information law and policy, especially copyright, free speech and privacy issues in schools and libraries. In fall of 2005, Professor Lipinski was placed on the Fulbright Senior Specialist Roster and was named a member of the Global Law Faculty, University of Leuven in Fall of 2006. Email: lipinski@sois.uwm.edu

Daniela Luzi is researcher of the National Research Council at the Institute of research on populations and social politics. Her interest in Grey Literature started

at the Italian national reference centre for SIGLE at the beginning of her career and continued carrying out research on GL databases, electronic information and open archives. She has always attended the International GL conferences and in 2000 she obtained an award for outstanding achievement in the field of grey literature by the Literati Club. Email: d.luzi@irpps.cnr.it

Bertrum H. MacDonald is the Associate Dean (Research) in the Faculty of Management at Dalhousie University, Halifax, Canada. He holds a BSc (Biology) from Acadia University, and an MA (History of Science), MLS, and PhD (Information Science) from the University of Western Ontario. He is the Editor of Electronic Resources for the History of the Book in Canada / Histoire du livre et de l'imprimé au Canada project www.hbic.library.utoronto.ca>, funded by a $2.3 million grant from the Social Sciences and Humanities Research Council of Canada. His research focuses primarily on the dissemination of scientific information among Canadian scientists and engineers (both historically and currently) and on the history of scientific and technical publishing in Canada. In 2001 he was named a Dibner Library Research Scholar at the Smithsonian Institution in Washington DC. In recognition of his significant contributions to the bibliography of Canadian science and technology he was awarded the Marie Tremaine Medal by the Bibliographical Society of Canada in 2000 and is the GreyNet Award Recipient in 2004. Email: bertrum.macdonald@dal.ca

Elizabeth Newbold - After completing a degree in applied sciences and an MSc in Information Studies Elizabeth started a career as an information professional specialising in scientific, medical and technical information provision. Elizabeth has worked in a number of organisations and specialist information units, providing information research services to researchers in both the public and private sector before moving to the British Library in 2003 to manage the science, technology and medicine (STM) collections. Currently leading the STM Content & Collections team for the British Library her work focuses on collection and content development strategies and policies. Her interest in grey literature has developed over the years from her experiences working in organisations that were both users and producers of grey literature. Email: Elizabeth.Newbold@bl.uk

Debbie Rabina is a Assistant Professor at Pratt Institute, School of Information and Library Science. Her areas of teaching and research include scholarly communication, LIS education, government and NGO information sources, and information policy. Email: drabina@pratt.edu

Hans E. Roosendaal is Professor of scientific information at the Faculty of Philosophy and Social Science and the Faculty of Computer Science at the University of Twente in the Netherlands. Before, he has been member of the Executive Board and Director of Scientific Information at the University of Twente. Educated as physicist, Hans joined the University of Bielefeld (Germany) as faculty staff in 1974. Between 1983 and 1998 he served Elsevier Science in various management positions as a publisher and in corporate strategy and acquisitions. Since 1998 he is at the University of Twente. Hans authored about 50 articles and co-authored a

book in surface physics and he authored a number of articles on scientific information, in particular on strategic aspects of the transformation from a paper to a digital environment. Email: H.E.Roosendaal@utwente.nl

Joachim Schöpfel obtained his Ph.D. in psychology from the Hamburg University in 1992. From 1991 to 2008, he worked at the French Institute for Scientific and Technical Information (INIST-CNRS) in different positions in database production and library management, at last as head of the e-publishing and document supply department. During the same time, he was lecturer at the University of Nancy. At present, he is senior lecturer in information and communication sciences at the Charles de Gaulle University of Lille 3. He published on GL, document delivery, digital libraries, scientific publishing, usage statistics and professional development. Email: joachim.schopfel@univ-lille3.fr

Gretta E. Siegel has worked as a librarian since 1984 at universities, research institutes, and with state and federal agencies and Indian tribes. Her interests center on issues concerning access to information in the sciences and include topics such as the economics of scholarly communication, access to grey literature, subject mapping, and information literacy for graduate students. She has also made several trips to Cuba to observe and participate in their informaticization efforts. The author currently works at Portland State University in Portland, Oregon as their Science Librarian, Coordinator of Graduate Student Services, and Coordinator of Scholarly Communication Outreach and Initiatives. The author holds degrees in chemistry, biochemistry and biophysics, and information science. Email: siegelg@pdx.edu

Christiane Stock is the Head of the Monographs and Grey Literature service at INIST, in charge of the repositories LARA (reports), mémSIC (master's theses in information sciences) and OpenSIGLE. Member of the Technical Committee for the SIGLE database from 1993 to 2005, she also set up the national agency for ISRN (International Standard Report Number). She is member of the AFNOR expert group who prepared the recommended metadata scheme for French electronic theses (TEF). Email: christiane.stock@inist.fr

Markus Weber studied Sociology, English Linguistics and Social and Preventive Medicine at Zurich and Lausanne University in Switzerland. For many years he worked for institutions active in harm reduction measures (street work, shelter homes, etc.). On finishing his studies he started work at the Swiss Federal Office of Public Health in the Competence Centre for Evaluation (CCE) where he is responsible for designing, mandating, assessing and valorizing external evaluation studies on the Office's public measures (prevention, control, information). Email: markus.weber@bag.admin.ch

Peter G. Wells has recently retired from Environment Canada, though he is still there as a volunteer research scientist. His main affiliations are Professor and Adjunct Professor, Marine Affairs and Environmental and Resource Studies, Dalhousie University, Halifax, NS., and Adjunct Professor, Environmental Programs

and the Acadia Centre for Estaurine Research, Acadia University, Wolfville, NS. He has been very involved in a recent study on the grey literature of the Gulf of Maine Council on the Marine Environment (GOMC), comparing it to the previous study of GESAMP literature. The primary focus of the research team at Dalhousie is to examine the output, distribution, use and influence in policy and decision making of the grey literature of marine environmental intergovernmental groups, particularly the influence as the literature is expanding rapidly, while the ecosystem health and biological diversity of coastal seas and the oceans in many parts of the world are declining even faster. Email: peter.wells@ec.gc.ca

Appendix II

Index to Web based Resources in Grey Literature

GreySource provides examples of grey literature to the average net-user and in so doing profiles organizations responsible for its production and/or processing. Only web-based resources that explicitly refer to the term grey literature (or its equivalent in any language) are listed. GreySource identifies the hyperlink directly embedded in a resource, thus allowing immediate and virtual exposure to grey literature. The web-based resources appear within categories derived from the COSATI (American) and SIGLE (European) Classification Systems. The few changes that have been introduced into the classification scheme are intended to facilitate search and retrieval by net-users. (Date of Access, March 2010).

CLASSIFICATION SCHEME:

00 - GENERAL, MULTIDISCIPLINARY
01 - AERONAUTICS
02 - AGRICULTURE, FORESTRY, FISHERIES, VETERINARY SCIENCES
03 - ENVIRONMENTAL POLLUTION, PROTECTION AND CONTROL
04 - HUMANITIES (HISTORY, PHILOSOPHY, RELIGION, ETC.)
05 - SOCIAL SCIENCES (ECONOMICS, INFORMATION SCIENCE, ETC.)
06 - BIOLOGICAL & MEDICAL SCIENCES
07 - CHEMISTRY
08 - EARTH AND ATMOSPHERIC SCIENCES
09 - ELECTRONICS, ELECTRICAL ENGINEERING, COMPUTER SCIENCE
10 - ENERGY & POWER
11 - MATERIALS
12 - MATHEMATICAL SCIENCES
13 - MECHANICAL, INDUSTRIAL, CIVIL & MARINE ENGINEERING
14 - METHODS & EQUIPMENT
15 - MILITARY SCIENCES
16 - MISSILE TECHNOLOGY
17 - NAVIGATION, COMMUNICATION, DETECTION, ETC.
18 - SCIENCE AND TECHNOLOGY - (MULTIDISCIPLINARY)
19 - ORDNANCE
20 - PHYSICS
21 - PROPULSION & FUELS
22 - SPACE TECHNOLOGY

00 - GENERAL, MULTIDISCIPLINARY

Bibliotheksservice-Zentrum Baden-Wurttemberg
http://www2.bsz-bw.de/cms/recherche/links/fabio/fabioGRAU.html

BLDSC - British Library Document Supply Centre
http://www.bl.uk/reshelp/atyourdesk/docsupply/collection/rct/

J. Conrad Dunagan Library: Grey [or Gray] Literature
http://library.utpb.edu/greylit.html

EastView Information Services
http://www.eastview.com/russian/books/grey_literature.asp

GLISC, Grey Literature International Steering Committee
http://www.glisc.info

GreyNet, Grey literature Network Service
http://www.greynet.org

Grijze Literatuur in Nederland – GLIN
http://www.publiekwijzer.nl/bestanden.php?id=zoeknaar&db=3.2

Italian Grey Literature Database
http://www.bice.rm.cnr.it/letteratura_grigia_inglese.htm

LARA — Libre accès aux rapports scientifiques et techniques
http://lara.inist.fr/lara.jsp

Library Association of the City University of New York
http://lacuny.cuny.edu/committees/eis/fall2001/greyinvisible.html

OpenSIGLE - System for Information on Grey Literature in Europe
http://opensigle.inist.fr

02 - AGRICULTURE, FORESTRY, FISHERIES, VETERINARY ETC.

NAFRI, National Agriculture and Forestry Research Institute
http://www.nafri.org.la/03_information/greyliterature.htm

NCSU Natural Resources Library
http://www.lib.ncsu.edu/nrl/graylit.html

Pacific Fisheries Environmental Laboratory
http://www.pfeg.noaa.gov/research/publications/greyliterature.html

Pacific Regional Aquaculture Information Service for Education
http://praise.manoa.hawaii.edu/grayweb.php

Wildlands CPR
http://www.wildlandscpr.org/bibliographic-database-search

03 - ENVIRONMENTAL POLLUTION, PROTECTION AND CONTROL

Accessing Grey Literature of the Polar Regions
http://classic.ipy.org/development/eoi/details.php?id=162

BC Environmental and Occupational Health Research Network
http://www.bceohrn.ca/search/greylit/org

IMPROVE - Interagency Monitoring of Protected Visual Environments
http://vista.cira.colostate.edu/improve/Publications/GrayLit/gray_literature.htm

New Jersey Environmental Digital Library
http://njedl.rutgers.edu/njdlib/

04 - HUMANITIES (HISTORY, PHILOSOPHY, RELIGION, ETC.)

EURISLAM bibliograhic database
http://www.eurislam.info/index_EN.html

Touro College
http://www.touro.edu/library/GrayLit/GrayLiterature.asp

05 - SOCIAL SCIENCE, ECONOMICS, INFORMATION SCIENCE, ETC.

AIP, Archaeological Investigations Project
http://csweb.bournemouth.ac.uk/aip/aipintro.htm

Canadian Evaluation Society
http://www.evaluationcanada.ca/site.cgi?s=6&ss=8&_lang=an

COS West en Midden Brabant
http://www.cosnederland.nl/detail_proj.phtml?act_id=273&id=WMB&text03_tmp=WMB&text03=WMB

Criminology Library Grey Literature, University of Toronto
http://link.library.utoronto.ca/criminology/crimdoc/index.cfm

Documentation sur la Région des Grands Lacs Africains
http://www.grandslacs.net/home.html

ERIC - Education Resources Information Center
http://www.eric.ed.gov/ERICWebPortal/Home.portal?_nfpb=true&_pageLabel=NonJournalProvidersPage&logoutLink=false

GreyNet Conference Based Collections
http://opensigle.inist.fr/handle/10068/697753

Groningen State University, Library of Behavioural Social Sciences
http://www.rug.nl/bibliotheek/collecties/bibsocwet/grijzeliteratuur?lang=en

Haliburton County Collection
http://www.haliburtoncooperative.on.ca/literature/index.html

IMLS Grey Literature/DSpace Project
http://docushare.lib.rochester.edu/docushare/dsweb/View/Collection-331

Information for Practice
http://www.nyu.edu/socialwork/ip/

IZI, International Central Institute for Youth and Educational Television
http://www.izi-datenbank.de/en/

LAOAP, Latin American Open Archives Portal
http://lanic.utexas.edu/project/laoap/

National Archeological Database
http://www.cast.uark.edu/other/nps/nadb/nadb.mul.html

Milwaukee-based Public Policy Forum
http://milwaukeetalkie.blogspot.com/2007/12/were-gray-here-at-forum.html

National Library of Australia; Staff Papers
http://www.nla.gov.au/nla/staffpaper/amckenzie1.html

Online Bibliography of Anime and Manga Research
http://corneredangel.com/amwess/

PADI, Preserving Access to Digital Information
http://www.nla.gov.au/padi/topics/372.html

PsycEXTRA, a gray literature database
http://www.apa.org/psycextra/

Slaw, a co-operative web log about Canadian legal research and IT
http://www.slaw.ca/category/theme-grey-lit/

University of Central England in Birmingham
http://library.uce.ac.uk/edgreylitres.htm

University of New England, Learning Module
http://www.une.edu.au/library/eskillsplus/research/grey.php

06 - BIOLOGICAL & MEDICAL SCIENCES

BC Environmental and Occupational Health Research Network
http://www.bceohrn.ca/search/greylit/org

BELIT Bioethics Literature Database
http://library.wustl.edu/databases/about/belit.html

British Lichen Society
http://www.thebls.org.uk/content/survey.html

CADTH, Canadian Agency for Drugs and Technologies in Health
http://www.cadth.ca/index.php/en/cadth/products/grey-matters

Cochrane Reviews
http://www.cochrane.org/reviews/en/mr000010.html

ETH Zurich: Plant Pathology
http://www.path.ethz.ch/docs/grey

Fade: The North West Grey Literature Service
http://www.fade.nhs.uk

Health Technology Assessment (HTA) Information Resources
http://www.nlm.nih.gov/archive//2060905/nichsr/ehta/chapter10.html

Grey Literature Producing Organizations - New York Academy of Medicine
http://www.nyam.org/library/pages/grey_literature_producing_organizations

Grey Literature Report - New York Academy of Medicine
http://www.nyam.org/library/pages/grey_literature_report

Ornithological Worldwide Literature
http://www.birdlit.org/OWL

Searching for grey literature in medicine
http://blog.openmedicine.ca/node/253

Social Policy and Practice
http://bathhealthnews.blogspot.com/2009/11/new-database-social-policy-practice.html

The Survey, Women's Health Resources
http://thesurvey.womenshealthdata.ca/

University of Calgary - Health Science Library
http://libguides.ucalgary.ca/greylit

University of Waterloo
http://www.lib.uwaterloo.ca/discipline/health_kin/grey_literature.html

08 - EARTH AND ATMOSPHERIC SCIENCES

Bibliography of Chesapeake Bay Grey Literature
http://www.vims.edu/GreyLit/

CEDA, Centre for Environmental Data Archival
http://cedadocs.badc.rl.ac.uk/

Maryland Department of Natural Resources
http://www.dnr.state.md.us/irc/

09 – ELECTRONICS, ENGINEERING, COMPUTER SCIENCE

East European Technical Literature
http://www.tib.uni-hannover.de/en/special_collections/east_european/

10 - ENERGY & POWER

Environmental Science Research Guide - Grey Literature
http://libguides.acadiau.ca/content.php?pid=18724&sid=136803

ETDE, Energy Technology Data Exchange
http://www.etde.org/edb/fulltext.html

INIS- International Nuclear Information System
http://www.iaea.org/inisnkm/inis/products/aboutdb.htm

13 - MECHANICAL, INDUSTRIAL, CIVIL & MARINE ENGINEERING

Coastal Gray Literature
https://scholarsbank.uoregon.edu/xmlui/handle/1794/3781

MAGiC, Managing Access to Grey Literature Collections
http://www.magic.ac.uk/index1.html

18 - SCIENCE AND TECHNOL0GY (Multidisciplinary)

Grey Literature Science Sites
http://personal.ecu.edu/cooninb/Greyliterature.htm

Institute for Scientific and Technical Information
http://international.inist.fr/article55.html

20 - PHYSICS

Electronic Grey Literature in Accelerator Science and Its Allied Subjects
http://library.cern.ch/HEPLW/12/papers/4/

Appendix III
List of Grey Literature Document Types

This list was first compiled in 2004 during a study on citation analysis and grey literature in which 72 document types were cited. Since then, this list has been maintained on GreyNet's website and further developed by the international grey literature community.

A

Announcements
Annuals

B

Bibliographies
Blogs
Booklets
Brochures
Bulletin Boards
Bulletins

C

Call for Papers
Case Studies
Catalogues
Chronicles
Codebooks
Conference Papers
Conference Posters
Conference Proceedings
Country Profiles
Course Materials

D

Databases
Datasets
Datasheets
Deposited Papers
Directories
Dissertations
Doctoral Theses

E

E-Prints
E-texts
Essays
ETDs
Exchange Agreements

F

Fact Sheets
Feasibility Studies
Flyers
Folders

G

Glossaries
Government Documents
Green Papers
Guidebooks

H

Handbooks
House Journals

I

Image Directories
Inaugural Lectures
Indexes
Internet Reviews
Interviews

J

Journals:
Grey Journals
In-house Journals
Journal Articles
Non-commercial Journals
Synopsis Journals

L

Leaflets
Lectures
Legal documents
Legislation

M

Manuals
Memoranda

N

Newsgroups
Newsletters
Notebooks

O

Off-prints
Orations

P

Pamphlets
Papers
Patents
Policy Documents
Policy Statements
Posters
Précis Articles
Preprints
Press Releases
Proceedings
Product Data
Programs
Project Information
Documents
Proposals

Q

Questionnaires

R

Readers
Registers
Reports:
- Activity Reports
- Annual Reports
- Bank Reports
- Business Reports
- Committee Reports
- Compliance Reports
- Country Reports
- Draft Reports
- Feasibility Reports
- Government Reports
- Intelligence Reports
- Internal Reports
- Official Reports
- Policy Reports
- Progress Reports
- Regulatory Reports
- Site Reports
- Stockbroker Reports
- Technical Reports

Reprints
Research Memoranda
Research Notes
Research Proposals
Research Registers
Research Reports
Reviews
Risk Analyses

S

Satellite Data
Scientific Protocols
Scientific Visualizations
Show cards
Software
Specifications
Speeches
Standards
State of the Art Reviews
Statistical Surveys
Statistics
Supplements
Survey Results
Syllabi

T

Technical Documentation
Technical Notes
Tenders
Theses
Timelines
Trade Directories
Translations
Treatises

W

Website Reviews
WebPages
Websites
White Books
White Papers
Working Documents
Working Papers

Y

Yearbooks

Appendix IV

Collections of Conference based Papers 1993-2010

GL1 – First International Conference on Grey Literature, Weinberg Report 2000 (GL'93). – RAI Amsterdam (NL), December 13-15, 1993.
http://opensigle.inist.fr/handle/10068/

GL2 - Second International Conference on Grey Literature, Grey Exploitations in the 21st Century (GL'95). – Catholic University of America, Washington D.C. (USA), November 2-3, 1995.
http://opensigle.inist.fr/handle/10068/698012

GL3 - Third International Conference on Grey Literature, Perspectives on the Design and Transfer of Scientific and Technical Information (GL'97). – Jean Monnet Building, Luxembourg, November 13-14, 1997.
http://opensigle.inist.fr/handle/10068/697932

GL4 - Fourth International Conference on Grey Literature, New Frontiers in Grey Literature (GL'99). – Kellogg Conference Center, Washington D.C. (USA), December October 4-5, 1999.
http://opensigle.inist.fr/handle/10068/697891

GL5 - Fifth International Conference on Grey Literature, Grey Matters in the World of Networked Information. – KNAW Amsterdam (NL), December 4-5, 2003.
http://opensigle.inist.fr/handle/10068/697754

GL6 - Sixth International Conference on Grey Literature, Work on Grey in Progress. - New York Academy of Medicine (USA), December 6-7, 2004.
http://opensigle.inist.fr/handle/10068/697756

GL7 - Seventh International Conference on Grey Literature, Open Access to Grey Resources. – INIST/CNRS, Nancy (FR), December 5-6, 2005.
http://opensigle.inist.fr/handle/10068/697757

GL8 - Eighth International Conference on Grey Literature, Harnessing the Power of Grey. - University of New Orleans (USA), December 4-5, 2006.
http://opensigle.inist.fr/handle/10068/697758

GL9 - Ninth International Conference on Grey Literature, Grey Foundations in Information Landscape. - Provincial House Antwerp (BE), December 10-11, 2007.
http://opensigle.inist.fr/handle/10068/697759

GL10 - Tenth International Conference on Grey Literature, Designing the Grey Grid for Information Society. - Science Park Amsterdam (NL), December 8-9, 2008.
http://opensigle.inist.fr/handle/10068/697786

GL11 – Eleventh International Conference on Grey Literature, The Grey Mosaic, Piecing It All Together. – Library of Congress, Washington D.C. (USA), December 14-15, 2009.
http://opensigle.inist.fr/handle/10068/

GL12 - Twelfth International Conference on Grey Literature, Transparency in Grey Literature: Grey Tech Approaches to High Tech Issues. - National Technical Library, Prague (CZ) Forthcoming, December 6-7, 2010.

Appendix V

Thematic Index – The Grey Journal An International Journal on Grey Literature 2005-2010

Volume 1, Number 1, Spring	2005	Publish Grey or Perish
Volume 1, Number 2, Summer	2005	Repositories - Home2Grey
Volume 1, Number 3, Autumn	2005	Grey Areas in Education

Volume 2, Number 1, Spring	2006	Grey Matters for OAI
Volume 2, Number 2, Summer	2006	Collections on a Grey Scale
Volume 2, Number 3, Autumn	2006	Using Grey to Sustain Innovation

Volume 3, Number 1, Spring	2007	Grey Standards in Transition and Use
Volume 3, Number 2, Summer	2007	Academic and Scholarly Grey
Volume 3, Number 3, Autumn	2007	Mapping Grey Resources

Volume 4, Number 1, Spring	2008	Praxis and Theory in Grey Literature
Volume 4, Number 2, Summer	2008	Access to Grey in a Web Environment
Volume 4, Number 3, Autumn	2008	Making Grey more Visible

Volume 5, Number 1, Spring	2009	Paperless Initiatives for Grey Literature
Volume 5, Number 2, Summer	2009	Archaeology and Grey Literature
Volume 5, Number 3, Autumn	2009	Trusted Grey Sources and Resources

Volume 6, Number 1, Spring	2010	Government Alliance to Grey Literature
Volume 6, Number 2, Summer	2010	…
Volume 6, Number 3, Autumn	2010	…

Keyword Index