

UH



Universität Hamburg

Andrew Lass  
Richard E. Quandt  
(eds.)

# Union Catalogs at the Crossroad

Hamburg University Press



**Union Catalogs at the Crossroad**



# **Union Catalogs at the Crossroad**

Edited by  
Andrew Lass  
and  
Richard E. Quandt

Hamburg University Press  
Hamburg, Germany

## Impressum

Bibliographic information published by Die Deutsche Bibliothek

Die Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the Internet at <http://dnb.ddb.de>.

Bibliografische Information Der Deutschen Bibliothek

Die Deutsche Bibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.ddb.de> abrufbar.

This publication is also openly accessible at the publisher's website. Die Deutsche Bibliothek has archived the electronic publication, which is now permanently available on the archive server of Die Deutsche Bibliothek.

Diese Publikation ist auf der Verlagswebsite ebenfalls *open access* verfügbar. Die Deutsche Bibliothek hat die Netzpublikation archiviert. Diese ist dauerhaft auf dem Archivserver Der Deutschen Bibliothek verfügbar.

Available open access / *open access* verfügbar:

Hamburg University Press / Hamburg University Press

<http://hup.rrz.uni-hamburg.de>

Die Deutsche Bibliothek archive server / Archivserver Der Deutschen Bibliothek

<http://deposit.ddb.de/>

ISBN 3-937816-08-9

© 2004 Hamburg University Press, Hamburg

Rechtsträger: Universität Hamburg, Deutschland

Produktion: Elbe-Werkstätten GmbH, Hamburg, Deutschland

<http://www.ew-gmbh.de>

The Andrew W. Mellon Foundation and the National Library of Estonia organized a Conference on Union Catalogs which took place in Tallinn, in the National Library of Estonia on October 17–19, 2002. The Conference presented and discussed analytical papers dealing with various aspects of designing and implementing union catalogs and shared cataloging systems as revealed through the experiences of Eastern European, Baltic and South African research libraries. Here you can find the texts of the conference papers and the list of contributors and participants.



## Table of Contents

Union Catalogs in a Changing Library World: An Introduction .....	xi
<i>Andrew Lass and Richard E. Quandt</i>	

### Part 1 Western Models and Overview

Chapter 1.....	31
EUCAT: A Pan-European Index of Union Catalogs <i>Janifer Gatenby and Rein van Charldorp</i>	
Chapter 2 .....	51
The Virtual Union Catalog <i>Karen Coyle</i>	
Chapter 3 .....	67
The Cathedral and the Bazaar, Revisited: Union Catalogs and Federated WWW Information Services <i>Stefan Gradmann</i>	
Chapter 4 .....	89
Linking in Union Catalogs <i>Ole Husby</i>	
Chapter 5.....	101
Linda: The Union Catalog for Finnish Academic and Research Libraries <i>Annu Jauhiainen</i>	

Chapter 6 .....	123
Beyond Technology: Power and Culture in the Establishment of National Union Catalogs <i>Nadia Caidi</i>	

## Part 2 Czech and Slovak Union Catalogs

Chapter 7 .....	141
The CASLIN Union Catalog <i>Gabriela Krčmařová and Ilona Trtíková</i>	

Chapter 8 .....	173
LINCA: The Union Catalog of the Czech Academy of Sciences <i>Martin Lhoták</i>	

Chapter 9 .....	187
CASLIN Uniform Information Gateway <i>Bohdana Stoklasová and Pavel Krbec</i>	

Chapter 10 .....	205
The Slovak Union Catalog for Serials <i>Lýdia Sedláčková and Alojz Androvič</i>	

## Part 3 Polish Union Catalogs

Chapter 11 .....	227
Are Our Union Catalogs Satisfying Users' Needs? <i>Błażej Feret</i>	

Chapter 12 ..... 245

Union Catalogs for Poets

*Henryk Hollender*

Chapter 13..... 265

Aiming at the Union Catalog of Polish Libraries

*Anna Paluszkiewicz and Andrzej Padziński*

Chapter 14 ..... 281

Implementing KaRo: The Distributed Catalog of Polish Libraries

*Tomasz Wolniewicz*

#### Part 4 Hungarian Union Catalogs

Chapter 15 ..... 297

The Hungarian Shared Cataloging Project: MOKKA

*Géza Bakonyi*

Chapter 16 ..... 305

Subject Cataloging in a Cooperative Cataloging Environment

*Klára Koltay*

Chapter 17 ..... 327

Principles of a National Union Catalog: Shared Cataloging in  
a Small Country

*Erik I. Vajda*

Part 5 Baltic Union Catalogs

Chapter 18 ..... 341  
Using a Shared Cataloging System: The Estonian Approach  
*Janne Andresoo and Riin Olonen*

Part 6 South African Union Catalogs

Chapter 19 ..... 361  
A National Union Catalog for Shared Cataloging and  
Resource Sharing by Southern African Libraries  
*Pierre Malan*

Chapter 20 ..... 381  
Regional vs. National Union Database Development:  
The GAELIC Perspective  
*D. L. Man and Lettie Erasmus*

Chapter 21 ..... 407  
Why the “Big Bang” Did Not Happen: The CALICO Experience  
*Amanda Noble and Norma Read*

Contributors ..... 435

Conference Participants ..... 441

# **Union Catalogs in a Changing Library World: An Introduction**

Andrew Lass and Richard E. Quandt

## **1 The Background**

The papers in this volume were presented at a conference sponsored by The Andrew W. Mellon Foundation in Tallinn, October 17-19, 2002. The date of this conference was almost exactly on the fifth anniversary of another Mellon conference that took place in Warsaw and was devoted to library automation.

In 1989 and 1990, Hungary, Czechoslovakia and Poland abandoned their long-term obeisance to the Soviet system and began to chart a new course that would embrace democracy and market economies. It was quickly recognized in the West that the intellectual and financial restrictions under which these countries had operated would make the transition to a new political and economic system long and arduous, and western donors descended on these countries in droves to provide financial and technical assistance in democracy building, western style economics and modern management techniques. Notable government agencies in these efforts included USAID and USIA in the United States, and the European Bank for Reconstruction and Development, the PHARE program and the Tempus program in Europe; prominent among foundations were the Ford Foundation, the Pew Charitable Trusts, The Andrew W. Mellon Foundation, the Soros Foundation(s), the German Marshall Fund of the United States, and

numerous others.<sup>1</sup> In 1991, the Baltic countries achieved independence from the Soviet Union and became additional targets for western generosity.<sup>2</sup>

While numerous donors supported the development and modernization of the higher educational sector in these countries (and PHARE and Tempus were particularly noteworthy in this respect), relatively few donors realized either the crucial importance of research libraries to education and research or the extent to which ideology and financial stringency in the pre-1990 period had contributed to their inability to develop their collections and to keep up with modern western advances in library technology and user friendliness. The Mellon Foundation had accordingly decided to devote substantial resources to introducing modern western library automation technologies in the research libraries of the region.

The 1980s were not kind to authoritarian regimes. While the Communist system was experiencing strains as a result of Solidarity in Poland, the Civic Forum in Czechoslovakia and the increasing demands for independence in the Baltic countries, the system of apartheid in South Africa came under growing pressure from demonstrations, strikes and courageous academic leaders such as Stuart Saunders, the Vice Chancellor of the University of Cape Town (UCT). By February 1990, the Prime Minister, F. W. de Klerk, announced the removal of the ban on organizations such as the African National Council and the freeing of Nelson Mandela; by November of that year, Mandela could receive an honorary degree from UCT, and in 1994, general elections were held and Mandela was installed as President.<sup>3</sup> It was appropriate that the Mellon Foundation should, as did the Ford Foundation, the Soros Foundation and others, step into the breach

---

<sup>1</sup> See Richard E Quandt, *The Changing Landscape in Eastern Europe: A Personal Perspective on Philanthropy and Technology Transfer* (New York: Oxford University Press, 2002), chapter 2.

<sup>2</sup> Independence also came to the Balkan countries, but in comparison with their more northerly neighbors they appeared to be neglected by donors, if for no other reason than the disorderly situation that prevailed in the former Yugoslavia for a long time.

<sup>3</sup> Stuart Saunders, *Vice-Chancellor on a Tightrope* (Claremont, South Africa: David Philip Publishers, 2000).

and help repair the damage caused by the many years of apartheid. True to its general emphasis on and expertise in higher education, the Mellon Foundation took up this challenge, and libraries and their modernization constituted, as in Eastern Europe, one of the important foci of its activities.

By 1997, a substantial number of libraries in Eastern Europe had introduced western library automation techniques, and South Africa was on the verge of doing so. It seemed appropriate to pool the experiences of these transitional countries and to examine the ways in which they responded to their differing needs and circumstances, which was accomplished through the Warsaw conference in 1997.<sup>4</sup> Over a number of years, the libraries tried to overcome the Communist legacy of being merely passive repositories of knowledge and to become more user-friendly, cooperate with one another, and overturn the awkward institutional and organizational arrangements under which they were often forced to operate. Many libraries in Eastern Europe and South Africa formed consortia for the purpose of automation and were largely, if not immediately, successful in these efforts. The impetus toward operating consortially was largely economic—better terms from vendors, better utilization of manpower, greater unification of standards—but there were substantial differences in how consortia were implemented.

While it may have been premature in the late 1990s to declare victory in library automation, the fact is that much was accomplished, and the quality of the libraries in the various countries changed appreciably over the decade. One other consequence of consortial library automation was the discovery that it made excellent sense to think of library catalogs that covered not one library but an entire consortium of libraries; particularly one in which users could not only determine where in the consortium a particular item was held but what its borrowing status was at any one time. Focus turned to union catalogs, and it was only a small step from consortial union catalogs to national union catalogs. Thus came about the interest in a second Mellon conference, this time devoted to various implementations of union catalogs and related mechanisms.

---

<sup>4</sup> See Andrew Lass and Richard E. Quandt (eds.), *Library Automation in Transitional Societies: Lessons from Eastern Europe* (New York: Oxford University Press, 2000).

On October 16, 2002, some ninety-two librarians and information technology experts from thirteen different countries came together at the National Library of Estonia in Tallinn to share their experiences with building union catalogs and discuss a whole range of issues that inform their present strategies and future developments.

The international gathering was again funded by The Andrew W. Mellon Foundation and was organized by the Foundation in cooperation with the National Library of Estonia, which hosted it. The conference program included 34 papers, presented in eight panels over two days. Twenty-eight of these reported on Mellon-assisted projects in the Czech Republic, Slovakia, Hungary, Poland, South Africa, Estonia, and Latvia. Six specialists, from North America (USA and Canada) and Western Europe (Finland, Norway, Holland and Germany), shared their own experience with developing union catalogs and offered a more general perspective on some of the underlying technical and organizational issues. The morning of the third day was devoted to a lively panel discussion that was organized around the topics that had emerged as key, and even controversial, over the previous days.

As is the case with successful conferences, the possibility of hearing interesting and intriguing presentations was matched by an opportunity to talk to colleagues and discuss ideas and problems in detail. Add to this the wonderful ambience of the venue and the impeccable behind-the-scenes logistics of the hosts, and you have the makings of what turned out to be a both productive and memorable meeting.

The purpose of this volume is to present the wider audience of specialists with a selection of the papers presented at the conference. While all the presentations were very interesting, we decided to include in this volume those that we think best illustrate, in detailed case studies or retrospective analyses, the key problems facing the development of union catalogs in societies caught at the crossroad of two historically significant trajectories: the fundamental socio-political and economic transformations that are being experienced by these countries at a time when library and information services are themselves facing radical changes in their organization and mission, and in which the development of electronic technologies and the demands of globalization play a decisive role.

It is one of the ironies of the digital library age that many of the developments in the fields of science and technology have an increasingly

shorter half-life and traditional paper copy publications are increasingly expensive. It could be argued that it is not worth printing the proceedings from the conference on works in progress; after all, much will have changed by the time the book appears. We wish to argue an alternative perspective: it is precisely because any report on the status of union catalogs must be tentative and provisional that it is important for those facing the challenge of building union catalogs, as much as for those scholars interested in the history of library science, that we offer a report that captures this stage of development in printed form. A printed version is all the more important, since the archiving of purely electronic material is, if not in its infancy, not well developed, and standards are still being debated in the profession.<sup>5</sup>

## 2 Union Catalogs and the Mellon Foundation Initiative

Since 1997, all countries that have received library grants from the Mellon Foundation enabling them to implement an integrated library system have also started to work on or plan for union catalogs. The aim of the conference was to provide an opportunity to share their experience and compare the chosen methods and technologies with practices in other countries.

The experiences that the participants brought to the conference were wide-ranging and varied. In the Czech Republic, the Mellon Foundation supported union catalog efforts built on the CASLIN framework, which had implemented integrated library systems in a number of key libraries in the Czech Republic and Slovakia. In Slovakia, union catalogs were promoted by Foundation support for the National Library in Martin for retroconverting the Slovak National Bibliography, and by assisting the University Library of

---

<sup>5</sup> An important step toward standards is “Preserving Digital Information: Final Report and Recommendations,” a report by Donald Waters and John Garrett, under the auspices of the Commission on Preservation and Access and RLG. See <http://www.rlg.org/ArchTF/>. Also relevant here is the Mellon Foundation’s Ithaka Project, which has electronic archiving as one of its objectives.

Bratislava in its efforts to build a union catalog of periodical literature. In Poland, the Foundation funded NUKat, a centralized union catalog initiative based on the VIRTUA system of VTLIS, Inc. (A second and independent union catalog initiative, KaRo, is also implemented in Poland. However, it is likely that both efforts will benefit from their coexistence.) In Estonia and Latvia, employing the automation systems INNOPAC and ALEPH 500 respectively, the union catalog efforts stemmed directly from the consortial implementation of library automation. In South Africa, Mellon funded SABINET to replace the older SACat catalog with a technologically advanced national union database. Only in Hungary did union catalogs (VOCAL and MOKKA) come into being without direct Foundation assistance. More varied approaches are difficult to imagine, and we hoped that juxtaposing the experiences of such a varied group would provide instructive lessons.

### 3 Case Studies

Most papers presented at the conference were essentially case studies that provided an overview of specific union catalog projects. Some focused on the implementation of technologies aimed at introducing specific functionalities, while others chose to introduce the system in place, often against a historical background, and highlighted the problems encountered along the way. All projects addressed specific needs in untraditional ways as they juggled to make creative use of new technologies in a radically changed library information environment, and do so under a variety of real constraints (budgetary, legislative and organizational). And, as was to be expected, opinions differed on a whole variety of themes as much as the individual project strategies differed from each other. This diversity of approaches underlines the extent to which the concept of the union catalog has changed, a point well illustrated by the broad spectrum of answers offered by the panelists when asked, on the last day, to suggest a definition of the union catalog.

The Czech and Slovak Library Information Network (CASLIN), which has involved cooperation between several libraries in both countries since 1993, is represented in this volume by four papers. On the Czech side,

Stoklasová and Krbec discuss the cooperative effort between the National Library of the Czech Republic and Charles University (both in Prague) in developing and implementing a Web-based Uniform Information Gateway using SFX and MetaLib (Ex Libris). On the other hand, Krčmařová and Trtíková look at an effort, also at the National Library in Prague, to develop a centralized union catalog emphasizing the advantages of a locally developed system, CUBUS (based on Oracle), that recognizes international standards but caters to a heterogeneous environment in a cost-efficient way. While the UIG is designed with the end-user in mind, CUBUS was designed to empower technical services, particularly shared cataloging, among all the participating libraries. Finally, the trials and tribulations of developing a union catalog for the complex library system of the Czech Academy of Sciences, comprising 65 institutes, is the topic of Lhoták's paper. On the Slovak side, Sedláčková and Alojz Androvič describe the development of the Slovak Union catalog of periodicals, located at the University Library of Bratislava.

The situation in South Africa is covered by three papers, each devoted to one of the components of what amounts to an ambitious national library automation and union catalog project. The Western Cape library consortium (CALICO), consisting of four universities, is discussed in the paper by Reed and Noble. Theirs is a detailed discussion of the problems that were encountered along the way, which allows them to draw attention to the role that politics and human resource management play in projects that might be assumed, naively perhaps, to be dominated by mostly technical and economic hurdles. The perspective on developing a regional union database in the Gauteng consortium (GAELIC), comprising 16 separate institutions, is the topic of Man and Erasmus. While their paper focuses on the implementation of a shared cataloging protocol, it does so with reference to the adverse effect that the initial failure of the South African Bibliographical and Information Network had in the early efforts at library automation in South Africa. SABINET was to help introduce a proper cataloging protocol and develop a national union database. The full story of SABINET, established in 1983, and Sabinet Online, a new private company that took over SABINET's operational responsibilities, is the topic of a separate, detailed account by Malan.

The Estonian Library Network Consortium (ELNET) is the topic of Olonen and Andresoo's discussion. They provide a rare step-by-step description of the initial implementation process (INNOPAC being their system of choice), and turn their attention to the development of the shared union catalog ESTER that also functions as a national bibliography database.

Three Polish projects are discussed in this volume. Hollender offers a thoughtful meditation on the past and present vicissitudes of union catalogs. His discussion of the NUKat project (The National Universal Catalog of Poland) illustrates the challenges posed by the ever-present, but always changing, tension between the logic of cataloging and different search habits. NUKat is also the focus of Paluszkiewicz and Padziński. Their detailed discussion follows its development from the early stages, in which the focus was on the authority file, through the preparatory stages for the development of the actual catalog to its early stage of functioning. It concludes with an evaluation of the costs, as well as advantages, of the system in place. Finally, Wolniewicz explains the philosophy behind the recently launched Polish distributed library catalog KaRo (conceived as an alternative to NUKat). He discusses the functions, limitations and successes of this service, including some general observations about distributed services.

The Hungarian shared cataloging project (MOKKA), discussed by Bakonyi, once again illustrates the complexities of drawing together a large heterogeneous group of libraries (in this case 16) with different cataloging rules, five different integrated library systems, three different archiving formats, two different MARC formats, etc., into a fully functioning consortium. Koltay's paper focuses on subject access in the cooperative cataloging environment and uses as examples three cooperative databases in Hungary: the bibliographic databases of the Hungarian National Shared Catalog (MOKKA), the National Document Delivery System (ODR) and the Matriksz database (which itself consists of three subject heading systems used in Hungary, in addition to the UDC system). Vajda unveils the background and decision-making process that went into getting MOKKA off the ground in order to offer some interesting lessons for others to draw on, including the pros and cons of centralized and distributed union catalogs.

## 4 Functionalities

That technical and public services of libraries have faced a whole series of dramatic transformations in the new digital information age is, by now, a tired cliché.

The fact nevertheless remains that the union catalog has, as a consequence, moved to center stage of new library information systems. The traditional needs (such as shared cataloging, record quality control) or services (bibliographic searches, ILL) are now augmented by new ones: the possibility of online search and text delivery, single point of access, and a broader range of objects, including Internet sites, 2D (paintings, photographs) as well as 3D (museum) objects, sounds and moving images.

All of them raise questions about the appropriate description rules and linking standards, search engine algorithms, storage memory, licensing, user identity, and security, to name a few.

Within this ever-expanding and changing array of technological possibilities and implementation pitfalls, the final decision on the type of union catalog, its architecture, functionalities and, finally, vendor choice must lie with the libraries themselves. A thoughtful and step-by-step analysis of this decision-making process is the topic of Coyle's paper on the conversion of the University of California centralized union catalog MELVYL (that worked with broadcast searches of participating libraries) with a virtual catalog that could accomplish the same satisfactory results more efficiently, that is, both faster and at "a potential cost saving to the University."

Her discussion also highlights the one issue that comes up repeatedly: the relative advantages of distributed (virtual) and centralized (real) systems. While the virtual catalog could be said to be more current (in real time), it favors the more homogeneous environment (similar local systems, cataloging, indexing) and assumes that all systems are up at all times. The 'real' union catalogs are costlier, but have better control over record quality and operate independently of the participating institutions.

While the majority of case studies presented at the conference would fit on one or the other side of this dichotomy, some would argue that, in fact, the debate over the relative virtues of either type of architecture is somewhat misleading. For example, several of the papers (Gatenby and van

Charltdorp, Husby) make reference to the OAI (Open Archive Initiative) protocol, known also as ‘metadata harvesting,’ designed as “an application-independent interoperability framework”<sup>6</sup> that enables a union catalog to be maintained by libraries that operate different systems. Since this protocol enables libraries to run a union catalog in a heterogeneous environment without the use of standards (such as Z39.50), it also raises the possibility of operating them independently of the primary system vendors.

The uneasy relationship with vendors was, of course, one of the topics that came up several times during the conference, and the OAI protocol also illustrates the option of in-house development of union catalog modules that are tailored to specific needs. For example, the Oracle-based union catalog of the Czech National Library was designed locally, and was meant to supplement the main library system (ALEPH 500) and allow participation with libraries that could not afford the Z39.50 protocol license.

## 5 Links and Clicks

Perhaps the most significant development in the area of information delivery is the World Wide Web and its various search engines. The question becomes: what is the exact relationship between the Web-based information service and the electronic library (union) catalog? If information is organized differently in the two systems, what happens when information in one points to information in the other? To what extent can two different mechanisms for the organization of information coexist in what could be considered a hybrid setting? Gradmann’s paper takes on the task of identifying these differences “in terms of mutual redundancy, competition and (sometimes and hopefully) convergence”. While several of the papers actually identified the Web as the proper vehicle for the union catalog, it was also clear that preferences were very much linked to the primary purpose that the catalog was intended for. Those catalog projects that were focused on

---

<sup>6</sup> See <http://www.openarchives.org/OAI/openarchivesprotocol.html> for full discussion of the protocol.

traditional library needs and materials (such as bibliographic descriptions, copy cataloging, etc.) seemed less concerned with this issue than those that aim to provide the user with a single access point to ‘one-stop’ shopping for a range of types of information. Here the cooperation between Charles University and the Czech National Library, using the Open URL protocol (and MetaLib), is particularly interesting (Stoklasová and Krbec).

But, as Husby’s paper on linking in union catalogs points out, whether one is working with a Web-based or the more traditional electronic-based database (catalog) or, more precisely, because today one needs to work with both, the very concept of reference—the principal mechanism of any library information system providing the link between metadata and a specific object, or between objects—demands further clarification. As does the concept of holdings: among other things, network documents do not reside on library shelves and an increasing number of objects are complex, consisting of text in addition to other materials, themselves residing in different ‘locations.’

## 6 Costs and Benefits

It is fundamental in the design of capital improvements to consider the costs and benefits of the proposed changes. Only if the discounted value of the stream of future benefits exceeds the present value of costs could one argue rationally that the improvement should be carried out. This principle is, of course, a direct consequence of placing library decisions in an optimization framework and requiring that the decisions made satisfy some social optimality conditions.

While such calculations may not be easy, particularly because the stream of benefits is difficult to identify, let alone quantify, optimality calculations are typically not undertaken in the library context, not even in the simplest cases such as the question of purchasing the optimal number of software licenses when introducing an integrated automated library system. But optima have been determined in some such cases, and it would be extremely beneficial if librarians and those who control budgets would at

least be willing to think in these terms.<sup>7</sup> On the whole, it seems to be the case that those who work in the Anglo-American tradition, being perhaps more used to formal economic modeling, are somewhat likelier to think in terms of cost-benefit analysis. A notable exception to this generalization is the paper by Feret, who wants to determine the benefits that users derive from union catalogs. Malan's paper may be the only one that explicitly deals with costs and benefits due to shared cataloging, and contrasts the explicit costs of original and copy cataloging. Man and Erasmus pay significant attention to the financial benefits that accrued to libraries as a result of the GAELIC consortium, and note that cost savings arise from copying records from OCLC WorldCat. Read and Noble note that rising prices of print subscriptions have serious implications for library policy, while Jauhainen asks whether centralization of functions could save money. But for most authors, the discussion of costs and benefits is peripheral, and it is fair to say that the papers do not on the whole come to grips with these questions.

## 7 Cooperation

If there is a central underlying theme to most of the papers in this volume, then it is the importance of cooperation, whether intra- or inter-library and whether defined by consortial agreements are not, to the success of library automation project and most particularly to the building a union catalogs. From the very outset, libraries must agree on basic strategies (for example, whether to follow a distributed or centralized model), agree on standards, cost sharing, network strategies and, finally, the approach to the various technical and public services pursued. And of course it is not enough to agree; these agreements must be upheld, as library managements need to make a transition to strategies that allow their institutions to thrive in a separate but equal setting. And how does one judge the success of a union catalog project, given the multiplicity of players and factors involved? Two

---

<sup>7</sup> For a case in point, see Richard E. Quandt, "On the Optimum Number of Library Software Licenses," *Journal of Economic Behavior & Organization* 38/3 (1999): 349–56.

papers in this volume address this issue. Feret discusses the importance of establishing benchmarks for the evaluation of union catalog functionalities, including performance indicators that reflect user satisfaction for the development and running of union catalogs. He also suggests the appropriate methodologies for designing user satisfaction surveys. Caidi's paper presents the results of her survey of the Mellon-funded union catalog projects in Eastern Europe. This comparative study takes a closer look at the extent to which the development of national union catalogs was influenced by choices that were not technical. She makes the point that while the technologies used are globally available, their implementation is always local. Any library's vision (or "philosophy") of a union catalog is therefore informed by different social practices and cultural histories.

## 8 Politics

Many authors have made a point of highlighting the political dimensions of union catalog projects. In the most general sense, developing and maintaining a union catalog of any type rests, explicitly or not, on several social factors that may appear to lie outside the purely technical issues although, in fact, they are inseparable from them. As noted above, union catalogs are, by definition, built with the idea of cooperation between different libraries, even competing ones, on the continued support—financial, logistical and even legislative—of oversight organizations (e.g. universities, regional or national governments, different ministries) and, in no small manner, on the internal cooperation between the different parties that are directly involved in the functioning of the catalog (librarians, IT personnel, vendors and even users). In the end, any union catalog “emerges as a result of the interaction between these different players; it becomes an artifact that is socially constructed by people who have a stake in its development.” (Caidi) But even in cases where differences of opinion and personal agendas are a matter of organizational management, external (to the institution) political factors also play a decisive part.

The close ties between national political agendas and the direction that union catalogs pursue is, of course, the underlying theme of all the case studies. Two reasons stand out.

First, the very point of the Mellon funding was to assist with the development of UC projects in institutions that had not only just gone through the library automation challenge, but that had done so as part and parcel of the political transformation of the whole country (the fall of Communism and the end of apartheid). The type and condition of union catalog initiatives, to the extent that they existed, can be directly linked to the policies (and resources) of the previous regimes, and in several instances the new projects tried to work from these rather than start entirely new ones. It is important to keep in mind as well that the union catalog concept has a well-established historical precedent in all the countries represented here. The present projects' trajectories are therefore informed by the past, and often by a very conscious attempt to work with existing databases and established obligations, while introducing new standards or moving away from constraints that had political agendas and negative consequences. For example, the cooperation and division of labor between libraries in the Czech and Slovak case reflected the existence of one country. Up until 1993, the National Library in Prague (Bohemia) focused on a catalog of foreign literatures, while the University Library in Bratislava (Slovakia) focused on periodicals. The breakup of Czechoslovakia into two countries had a profound impact on how these national union catalogs were conceived and what form of cooperation, if any, would exist between the new, separate entities. Similarly, the end of apartheid in South Africa made it possible to bring existing but failing initiatives back to life, but also called for new and untested levels of cooperation between institutions previously separated by the racial divide.

Second, many of the initiatives—and this is particularly so in the case of European libraries—are located at the National Libraries and therefore are meant to fulfill their role as a central comprehensive service. In several cases, developing a national union catalog is mandated by law and may even require that all participating libraries operate under the same architecture (vendor). In other words, technical discussions regarding the relative merits or challenges of union catalogs operating with homogenous or heterogeneous environments may be decided by external political considerations. Compare, for example, the Slovak library legislation, which stipulates a unified system for all major libraries, i.e. single vendor, and the Czech legislation, which mandates the National Library to house the national

union catalog but allows for multiple systems. Ironically, as traditionally centralist systems try to give way to relative regional autonomies, national institutions such as national libraries become key players and lobbyists for regulations that can be perceived by other libraries as undermining a process that would support horizontal cooperation amongst administratively decentralized institutions.

University library UCs may be no less politicized by the nature of their relationship with the university's administration, such as a Dean in cases where the union catalog is meant to integrate individual departmental libraries within one school (e.g. School of Humanities of Charles University, Prague), or the Rector's office in all-university catalogs. Inter-university consortia pose their problems as well. The Polish example and two South African examples illustrate the potential hazards. The CALICO and GAELIC projects are particularly telling, as both of these try to integrate institutions that had minimal, if any, contact under apartheid rule (also a strong presence in the SABINET case). It could be argued that the degree of success of these politically 'heterogeneous' consortia is a direct reflection on their ability to develop and sustain social relations that transcend the dysfunctional, though well-entrenched, order.

Similarly, it would be interesting to speculate whether the degree of success of international consortia is a direct consequence of the relative stability of the institutions involved, the relation between the countries involved and the actual functionalities offered. For example, EUCAT, originally established in 1979 as a catalog linking both national and individual union catalogs in France, Germany and the Netherlands, is set up to grow and function as a pan-European index of union catalogs providing 'one-stop' access to full bibliographic searches with links to individual libraries (and ILL), document delivery services or links to full electronic texts.

## 9 Concluding Remarks

The past few decades have witnessed a revolutionary expansion in the functions, services and methodologies of libraries, and an equally remarkable growth in information resources that are no longer synonymous

with the traditional library. The traditional library today is only one of a multitude of information providers, and has had to adapt to, and indeed exploit, the availability of the World Wide Web. In the process of doing so, librarians have had to address many tough questions ranging from the user-friendliness of access to information to the proper role of union catalogs and the advantages or disadvantages of various ways of implementing them. The papers in the present volume amply illustrate the very substantial progress that has occurred, not only in technical accomplishments, but also in developing new modalities of cooperation in an environment in which it seems increasingly wrong-headed to strike out on one's own and in recognizing the 'political' dimension of problems that might have been naively thought to be purely technical. But we must end our introduction to this volume with a plea for more attention by librarians to a relatively neglected characteristic of providing access to information, namely the efficiency of the process and its costs and benefits. Libraries have the potential of providing rich data about their own operations that permit the application of techniques, usually developed in other contexts many years ago, for determining how efficiently a library operates and what the costs and benefits are of alternative ways of providing access to seekers of information. We have already alluded to one optimization model in a library context (see footnote 7). We mention here three more in the hope that the ever-present scarcity of resources will induce librarians to include economic analyses in their planning. A statistical study that relates the aggregate cost of various library services to the quantity of those services delivered is provided by Lewis G. Liu.<sup>8</sup> The well-known technique of frontier production functions, which employs econometric methodology to find the relationship between inputs and the maximum output that can be secured from them, is discussed in the context of museums by Bishop and Brand.<sup>9</sup> Data envelopment analysis, a technique based on linear programming

---

<sup>8</sup> Lewis G. Liu, "The Cost Function and Scale Economies in Academic Research Libraries," *Library Trends* 51/3 (2003): 293–311.

<sup>9</sup> P. Bishop and S. Brand, "The Efficiency of Museums: A Stochastic Frontier Production Function Approach," *Applied Economics* 35/17 (2003): 1853–58.

and originally developed by Charnes et al.,<sup>10</sup> is applied to libraries by Shim.<sup>11</sup> What all these studies have in common is that they apply formal mathematical or econometric techniques to evaluating library performance from the economic point of view. We hope that the application of such techniques will become as commonplace in library circles as the discussion of library and Internet technology.

---

<sup>10</sup> A. Charnes, W. W. Cooper, A. Y. Levin and L. M. Seiford, *Data Envelopment Analysis: Theory, Methodology and Application* (Boston: Kluwer, 1994).

<sup>11</sup> Wonsik Shim, "Applying DEA Technique to Library Evaluation in Academic Research Libraries," *Library Trends* 51/3 (2003): 312–32.



## Part 1

### Western Models and Overview



# Chapter 1

## EUCAT: A Pan-European Index of Union Catalogs

### Why a Pan-European Index?

Janifer Gatenby and Rein van Charldorp

End-users want a single, comprehensive, online source as exemplified by the success of Internet search engines, of which Google is a notable example. In just 4 years, Google has indexed more than 2 billion URLs and has grown to be the most-used search engine.<sup>1</sup> As a result of experience with such search engines, users are increasingly expressing a desire for a single point of access to library resources.

A single point of access to European library resources would offer end-users comprehensive, high quality, verified materials, with access to related materials, online text, and delivery services for offline materials.

From the user perspective, it is the content and comprehensive coverage that are important, far more important than the software, techniques and protocols used to achieve the interface and service. Moreover, users want access to content without having to learn the names and coverage of all the databases that would potentially house what they need. They want access with the minimum of training. Users do not have the same needs, nor does any one user always want the same type of information, at times requiring exhaustiveness and at others just what is readily available. What people require is a system that is flexible in the views that it can present.

---

<sup>1</sup> Google Inc. Fact Sheet, <http://www.google.com/press/facts.html>.

From a library perspective, a single point of access to European resources would offer the ability to provide a comprehensive view of available library materials Europe-wide and worldwide, backed by inter-library loans and document delivery services. This would ensure maximum exposure to their collections. The index should serve as the pivotal point for document delivery services with the library in a central role, leading to both digital and non-digital materials and the necessary requisites for access where appropriate. The index should also provide a tool for cooperative collection building and allow the library to be a member of more than one contributing union catalog.

It is important to create an environment and architecture that enables union catalogs to flourish, since they would be the main contributors to a pan-European index. The diversity of languages, cultures, cataloging rules, subject and name authorities and classifications, and other national and regional conventions make the task of a fully centralized union catalog almost impossible and inoperative. A federated approach is therefore necessary to ensure comprehensive coverage. Additional potential benefits from participation in a large central index include maximized ILL and document delivery services, efficiencies in creating quality shared linking services and enriched data mining services. The contents of the union catalog can be analyzed statistically in relation with other union catalogs providing, for example, information for EU projects and cooperative efforts in general.

To date, there have been numerous attempts to create virtual union catalogs, with Z39.50 as the key protocol in achieving this via broadcast searching. Examples of such virtual catalogs include the ONE project, the TEL project, the Canadian Virtual Union Catalog (VCUC) and the Texan Union catalog (ZLOT). These projects have all achieved moderate success, but the more individual catalogs that are searched simultaneously, generally the more slowly the results are presented, and retrieval from those catalogs is inconsistent due to differences in data indexing. Indexing differences also mean that searches that are common to all databases are few and basic, and consequently, precise broadcast searches are often not possible. Duplicates are retrieved in the results, and duplicate detection and grouping fail to work in a timely fashion over large result sets—just when they are most needed. These principal drawbacks to virtual catalogs have resulted in interest in the

harvesting model used by Internet search engines. If the data can be gathered, loaded and indexed centrally with duplicates removed or grouped as part of the update process, then the major drawbacks of virtual catalogs could be overcome. The OAI protocol, originally conceived for the harvesting of documents, was recently being considered for the creation of physical, non-virtual centralized indexes and union catalogs. Instead of the virtual catalog, a better model would be a centralized index for searching links to full text, to specific catalogs for services such as loans and photocopies, and to suppliers of other services, e.g. online book suppliers, databases of reviews, biographies, encyclopedic articles and so on.

Building such a centralized index could only be possible if done in a cooperative manner. Some tasks are so great they can only be achieved with the cooperation of a large number of parties, some of whom are otherwise competitors. Realizing a single comprehensive user point of access is critical to the continued central role of libraries in information provision. If libraries were to drop out of the limelight and their funding consequently reduced, it is likely that their role in cultural preservation would be difficult to fulfil. The world would enter a period of information chaos due to the concurrent upheaval in publishing, where it is easy to publish directly on the Web without peer review, control of document authenticity, or preservation and archiving. The result would be a permanent loss to the cultural heritage.

## 1 Description of EUCAT

EUCAT was conceived by OCLC PICA. As a not-for-profit company that is fully dedicated to libraries, it has served since the 1970s and supports major library installations in the Netherlands, Germany, and France. It is a company that has the business infrastructure, experience, software, and human resources to realize a pan-European index.

EUCAT is a pan-European index of union catalogs. It provides a quality catalog based on metadata, with duplicates identified and grouped and with authority control of authors, subjects and other headings ensuring consistent indexing and recall.

The main focus of EUCAT is as a discovery tool, linking to the contributing union catalogs or individual catalogs for services, in particular

inter-library loans and linking to licensed full text services. Document supply services, enriched contents, reviews, abstracts, and e-books may be accessed directly from EUCAT or from the participating union catalogs. EUCAT is the entry point for discovering and locating the riches of European libraries, both physical and digital.

EUCAT will also be used to support cooperative collection development by providing statistical analysis of coverage and by allowing libraries to record areas of proposed intensive development, e.g. digital projects. An additional benefit achieved by a centralized index is that in itself it becomes an authoritative source by its size alone and by being based on the resources of libraries, i.e. professionally created collections. This is particularly important at a time when the controls of traditional publication with editorial and peer review are being severely challenged by easy online publication and distribution. National libraries are among the major contributors to EUCAT that gives a combined index of national bibliographies and legal deposit indexes. EUCAT is thus a resource for establishing the authenticity of published and publicly available works.

Initially, EUCAT is not a source of copy cataloging. The index can direct to union catalogs from which copy cataloging may be made available depending on local arrangements. The index's main purpose is discovery; to add copy cataloging would entail complex arrangements to ensure the participation of some, and may deter some important libraries from participating.

OCLC PICA makes EUCAT available through different services, principally PiCarta and Publiekwijzer. It will also be possible to access EUCAT via external interfaces and portals using a search protocol, in particular ZING/SRU or SRW, Z39.50 and OpenURL.

## 2 Current Composition of EUCAT and Current Services

EUCAT currently consists of the holdings of the Dutch Union Catalog (*Nederlandse Centrale Catalogus*, NCC) and the libraries of the North German States (*Gemeinsamer Bibliotheksverbund*, GBV).

The central union catalog of the Netherlands represents the holdings of 642 libraries associated with 14.5 million bibliographic records. Text-based

materials, books, articles and serials represent nearly 95% of resources, with the remainder being printed music, sound, audio-visual and online resources (see Figure 1).

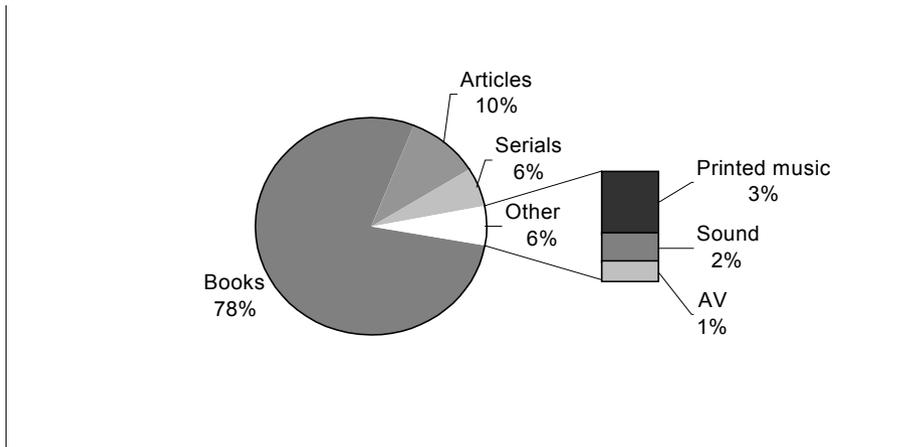


Figure 1. Form of Material Analysis of NCC

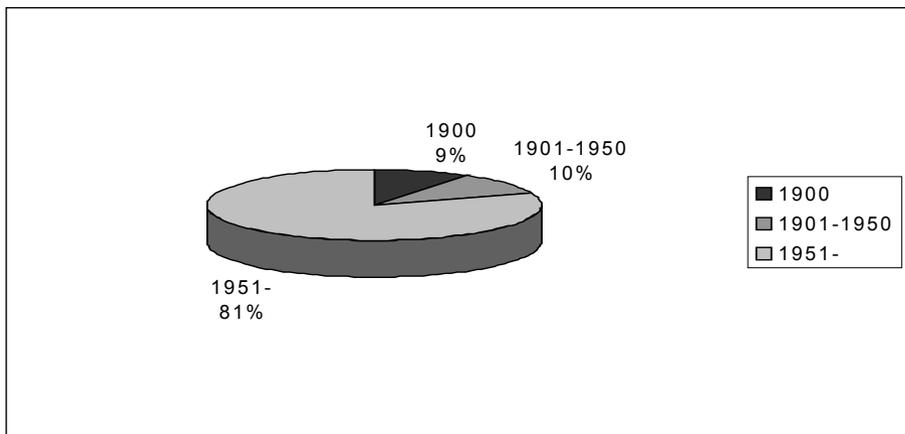


Figure 2. Date of Publication Analysis of NCC

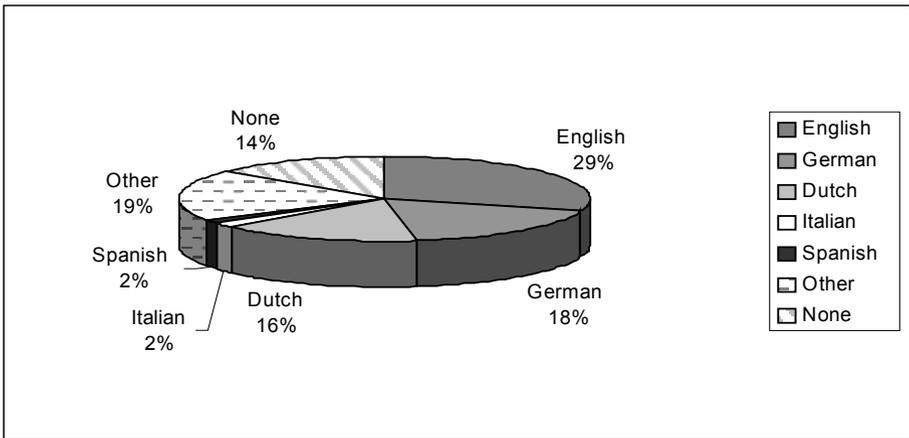


Figure 3. Language Analysis of NCC

Approximately 9% of titles were published before 1900, and over 75% since 1951 (see Figure 2).

The NCC catalog is available to Dutch end-users through the PiCarta service and to external systems via the Z39.50 protocol. As well as including EUCAT, the PiCarta service in the Netherlands also includes online contents data consisting of metadata and abstracts that are linked to full text services (depending on the license of the library). Users may view either the entire service or a specific catalog. Currently there are on average 150,000 inter-library loan requests and 40,000 copy requests per annum. There are just under 16,000 Dutch end-users who directly (i.e. unmediated) generate 40% of the loan requests and 20% of the copy requests. Direct access to full text (document delivery) is growing steadily, from just 5,000 in 2000 to an estimated 55,000 in 2002 (January to August showed 36,000). This is expected to continue growing, replacing inter-library loan requests. Figure 4 indicates end-user requests.

GBV records comprise the holdings of over 400 libraries, or 37 million holdings associated with 20 million bibliographic records. These have been loaded to EUCAT and matched with the Dutch bibliographic records. Where a match has occurred, a link is made between the records such that

the holdings of both records can be viewed no matter which record is retrieved and displayed.

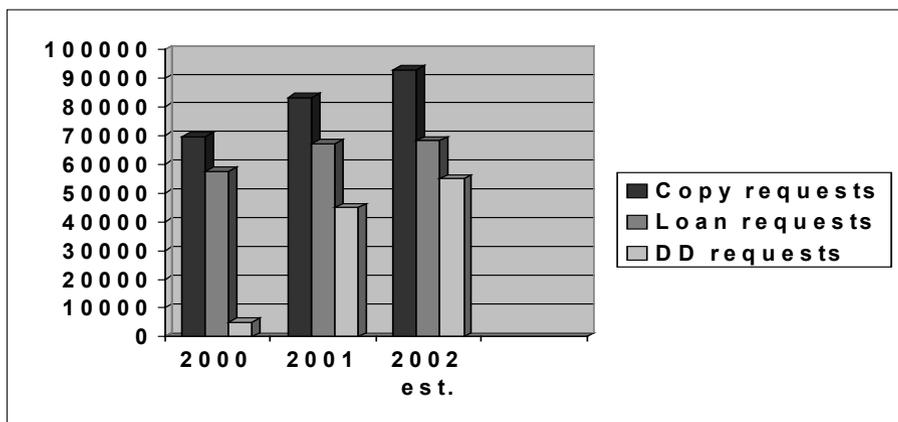


Figure 4. End-User Requests

The most appropriate record is displayed, depending on the user's login and reflecting the language of cataloging, subject headings and classification. This grouping and merging has the effect of virtual enhancement. For example, one record may contain a classification number not present in the other, but both records are accessible from the single point of access.

Approximately 20% of GBV records have been clustered with NCC records. The actual number of duplicates could be as high as 30% if algorithms as well as standard identifiers were used in matching.

Negotiations are underway for the users of GBV to have access to EUCAT through the PiCarta service. The logistics of international inter-library loans have not yet fully evolved because libraries are reluctant to give end-users direct access to this facility. Therefore, all requests are directed first to the user's union catalog for inter-library loans, and libraries will be able to restrict placement of international requests.

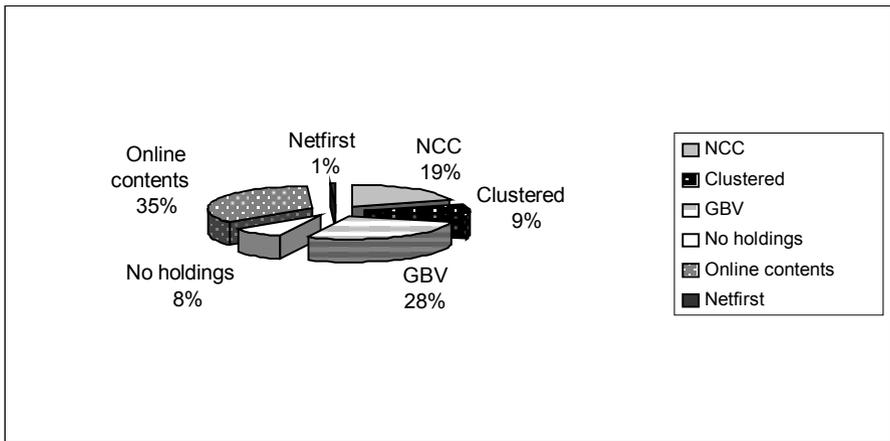


Figure 5. Database Composition of PiCarta

### 3 Expansion

The current contributors to EUCAT use the OCLC PICA CBS system for catalog maintenance and inter-library loans. This is not envisaged as a constraint on the system. Contributions from all major European union catalogs and libraries are necessary to provide the ideal index and single access point. As a first step in broadening the coverage of EUCAT, the European holdings from WorldCat will be loaded into EUCAT as a mirror copy.

OCLC WorldCat currently contains 21 million holdings from 430 European libraries. Actual figures on language, date coverage, and material type will be determined after loading. Arrangements for connecting to the various inter-library loan systems are currently being investigated, as is the determination of the business models required to ensure the widest possible cooperative participation in EUCAT. European WorldCat holdings represent full holdings for some libraries, and only the results of retrospective conversion projects for other libraries. The ideal situation would thus be to load from both union catalogs when possible, as well as from local catalogs if necessary.

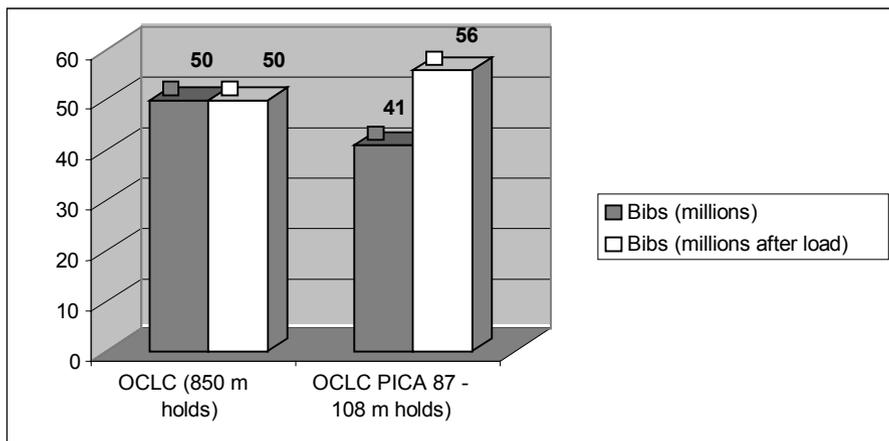


Figure 6. Estimated Unique Bibliographic Records

Figure 6 indicates the total bibliographic records of OCLC and compares them with the estimated total unique bibliographic records of all the OCLC PICA installations. It is estimated that the EU bibliographic records, with their holdings from OCLC's WorldCat, could add an additional 15 million unique bibliographic records to OCLC PICA's pool, and that the holdings would grow by 21 million, from 87 million to 108 million. Interestingly, it is estimated that fewer than 20% of Dutch titles and fewer than 40% of German titles in the Dutch national union catalog are also represented in OCLC's WorldCat. Similar overlap figures could be expected for other European national union catalogs, so it is clear that the bibliographic pool would be significantly increased with the contributions of such catalogs.

## 4 EUCAT Architecture and Standards

### Architecture

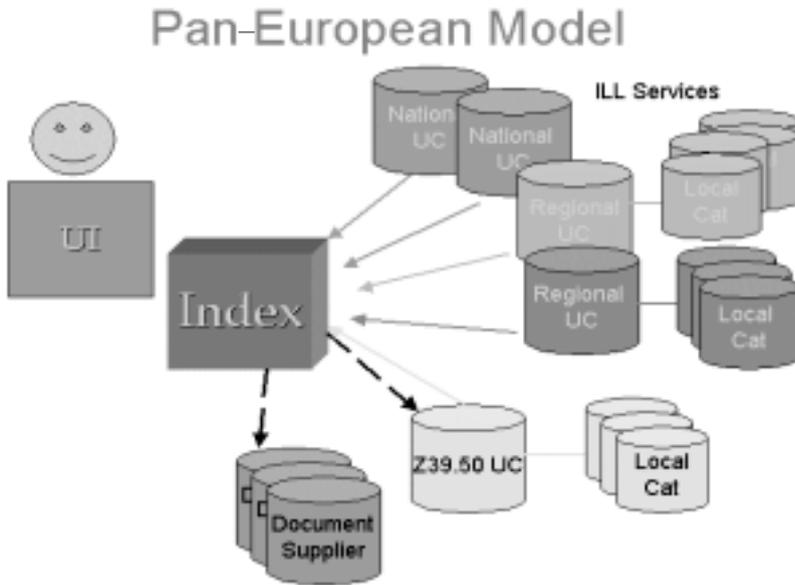


Figure 7. Pan-European Catalog

As new records are added and existing records are changed and deleted in the NCC and GBV union catalogs, they are also pushed directly (in the background) to EUCAT. All systems currently use the same OCLC PICA system, CBS/PSI, so that there is no need for an intermediary protocol. As others participate in EUCAT, alternative update mechanisms will be provided, most probably with a Z39.50 update (UCP profile) where the union catalogs push data to EUCAT, or with the OAI harvesting protocol, where EUCAT would poll the external systems, thus pulling the data. Consideration is being given to including the data structures of the UCP

with OAI, so that it will correctly handle modification and deletions as well as additions. Batch loading via FTP will also probably be an option. From the user's viewpoint, the index can show everything or be filtered regionally, by format, by language or by other criteria such as date. By default, the records are filtered to show local records first, together with local holdings, with access to other regional holdings.

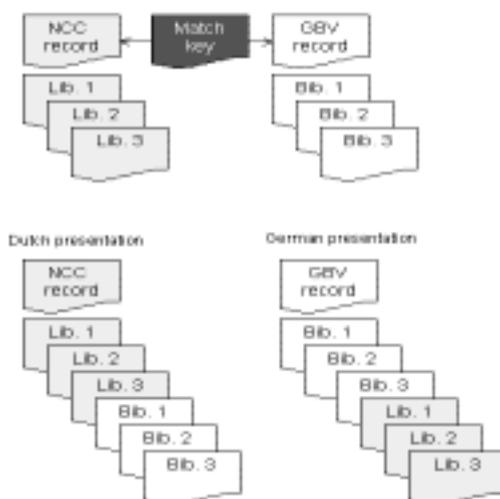


Figure 8. Record Filtering

End-user authentication is important both to determine the default views and confirm access to document delivery and other services. External authentication servers can be accessed using standards protocols such as LDAP and Athens.

EUCAT is also a part of WorldCat. EUCAT in Leiden and WorldCat in Dublin, Ohio will become the first two nodes of an extended WorldCat. Bibliographic and authority data from local nodes will be pooled together with a centralized and replicated international library directory. Holdings are held in the nodes, and hence the services are decentralized. A European and a world view will thus become available.

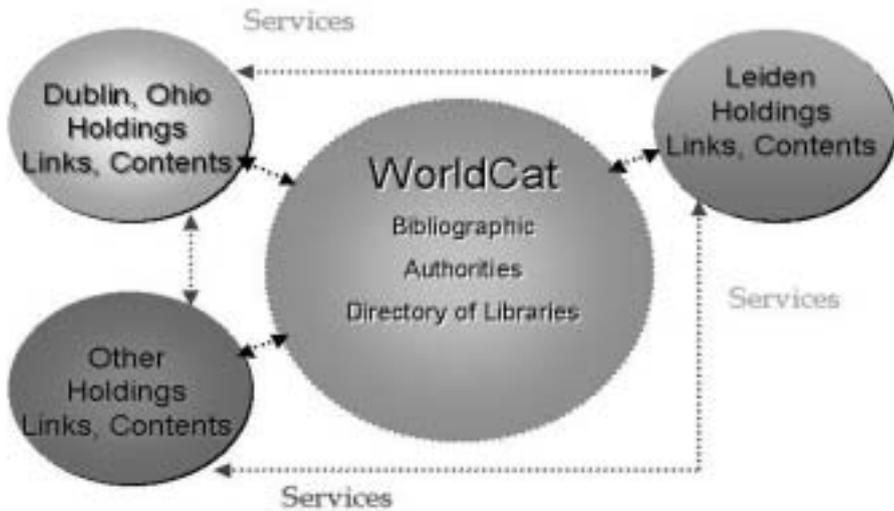


Figure 9. The Extended WorldCat Network

### Importance of Standards

To build a resource cooperatively and to ensure its general usability, standards are essential. EUCAT will be a resource shared by many different and disparate systems. It will have many different interfaces.

### Searching

So that external systems can access EUCAT for searching, the index will be available via standardized search protocols including:

Z39.50

(<http://www.loc.gov/z3950/agency/>) and the emerging

ZING SRW/ SRU

(<http://lcweb.loc.gov/z3950/agency/zing/srw/specifications.html>)

Bath Profile

(<http://www.nlc-bnc.ca/bath/>)

- Bibliographic
- Holdings
- Authorities
- Cross-domain

Other search standards may emerge.

These same protocol standards will be used to access those catalogs where the institutions have opted for partial participation in EUCAT.

## Updating

For updating the protocols, Z39.50 update (UCP), OAI and FTP have already been mentioned. One very important advantage of the EUCAT architecture is that it can accept and deliver multiple formats. The following syntaxes and schemas are possible:

### MARC

- ISO 2709: all variations including MARC21, UNIMARC
- MAB
- OAI XML encoding  
(<http://www.dlib.vt.edu/projects/OAi/marcxml/marcxml.html>)
- MODS (<http://www.loc.gov/standards/mods/>) .

### Metadata

- Dublin Core (<http://dublincore.org/>)
- ONIX (<http://www.editeur.org/onix.html>)
- METS (<http://www.loc.gov/standards/mets/>)

### Z39.50 Holdings schema

(<http://lcweb.loc.gov/z3950/agency/defns/holdings.html>).

FRBR is under investigation as a standard for the provision of better presentation and navigation. If it proves successful, then work is needed to incorporate FRBR elements into existing search and update protocols and to develop schemas for the structuring of records for exchange. FRBR also promises to permit copy cataloging at levels, and hence greater cataloging efficiency. (<http://www.ifla.org/VII/s13/frbr/frbr.pdf>)

## Other Standards

### OpenURL Standard

For linking, OpenURL is the emerging standard ([http://www.niso.org/committees/committee\\_ax.html](http://www.niso.org/committees/committee_ax.html)). On one level, this standard enables a simple identification number search to be run on a foreign server. It is used to discover full text, reviews, and related materials such as citation index materials, but also for order placement. What makes it different from other identifier standards like ISBN, ISSN etc. is that it is also a standard for the dynamic creation of an identifier for serial articles.

### ISO ILL (ISO 10160 / 10161)

(<http://www.nlc-bnc.ca/iso/ill/standard.htm>). This standard is currently undergoing minor revision.

### Circulation—NCIP

([http://www.niso.org/committees/committee\\_at.html](http://www.niso.org/committees/committee_at.html)). This new standard enables local systems to be accessed for the placement of loan requests and reservations and also to discover the status of items and users. It also includes authentication and is used as an alternative to more mainline authentication standards such as LDAP.

### Directories—ISO 2146

The ILL implementers' group (IPIG) is currently creating a structured library directory. This will be used as the basis for a revision of the ISO directory standard ISO 2146. Sections on curriculum strengths and reference services will be added to inter-library loan descriptive elements. The library directory will play an essential role in the extended WorldCat.

NISO is currently working on standards for an XML data schema and protocol for the exchange and forwarding of reference queries. ([http://www.niso.org/committees/committee\\_az.html](http://www.niso.org/committees/committee_az.html)). The main user interface to EUCAT will provide users with the ability to pose questions from any result page. Data from EUCAT can be used for creating a question or providing an answer.

## Cooperative Development and Experimentation

There is plenty of scope for cooperative development and experimentation among union catalogs that would be facilitated by participation in a common project. Examples are:

- Enhancements in retrieval; extension of the concepts of views and filters (by user, region, language, interest etc.);
- Improvement in the efficiencies of creating metadata—author-applied, program generation and extraction using algorithms, application of new information data models to evolve simple copy cataloging,
- Evolution in authority control to facilitate multilingual and multi-classification retrieval. FRANAR, VIAF and OCLC PICA's Colibri are project examples;
- Digital preservation and digital vault facilities;
- Sharing Web resources, pathfinders, predefined pages and links;
- Data mining to identify high-quality works that can be used in relevance ranking;
- Systems and programs for regular testing of URLs and 'shingles testing' to detect substantial changes in Web resources;
- Remote access, authentication and rights management.

## 5 Conclusion

EUCAT, as a centralized index to European library resources, provides fast, relevant and comprehensive searching through consistent indexing. Already large, it is capable of growing on a much larger scale, and OCLC PICA has the infrastructure to realize it. The cooperatively built index accommodates the diversity of European cataloging practices (codes, subject headings and classifications), languages and formats. It links to union catalogs, local catalogs and online providers for services. It is also capable of being linked from other services, e.g. e-learning environments, local library Web pages etc. and from abstract and indexing databases for European holdings.

EUCAT, with its association with WorldCat, can provide a gateway to library resources that rivals Google as a gateway to Internet resources.

Unlike Google, the library resource will lead to an available copy somewhere, the quality of resources retrieved is more consistent and the search capability yields more precise and complete results.

## Glossary and References

ABES	<i>L'Agence bibliographique de l'enseignement supérieur.</i> See <a href="http://www.sudoc.abes.fr">http://www.sudoc.abes.fr</a>
Bath Profile	The Bath Profile is an ISO Internationally Registered Profile (IRP) of the Z39.50 Information Retrieval Protocol, intended as a basis for effective interoperability between library and cross-domain applications. Conformance to this Profile's specifications will improve international or extra-national search and retrieval among library catalogs, union catalogs and other electronic resource discovery services worldwide. See <a href="http://www.nlc-bnc.ca/bath/ap-bath-e.htm">http://www.nlc-bnc.ca/bath/ap-bath-e.htm</a> .
DublinCore	Dublin Core Metadata Initiative. See <a href="http://dublincore.org/">http://dublincore.org/</a> .
EUCAT	EUCAT is a Pan European index of union catalogs. It may be accessed by the PiCarta service. See <a href="http://www.oclcpcica.org/?id=102&amp;ln=uk">http://www.oclcpcica.org/?id=102&amp;ln=uk</a> .
FRANAR	Functional Requirements of Authority Numbering and Records (FRANAR). The group is working on a conceptual model for authority information and international numbering for authority entities.
FRBR	Functional Requirements for Bibliographic Records. See <a href="http://www.ifla.org/VII/s13/frbr/frbr.htm">http://www.ifla.org/VII/s13/frbr/frbr.htm</a> .
FTP	File Transport Protocol. See <a href="http://searchnetworking.techtarget.com/sDefinition/0,,sid7_gci213976,00.html">http://searchnetworking.techtarget.com/sDefinition/0,,sid7_gci213976,00.html</a> .
GBV	<i>Gemeinsamer Bibliotheksverbund, GBV.</i> See <a href="http://www.gbv.de">http://www.gbv.de</a> .

ISO 2146	Documentation—Directories of libraries, archives, information and documentation centre, and their databases. See <a href="http://www.nla.gov.au/nla/staffpaper/jpearce2.html">http://www.nla.gov.au/nla/staffpaper/jpearce2.html</a> .
ISO 2709	Information and documentation—Format for Information Exchange. See <a href="http://www.iso.org/iso/en/CatalogDetailPage.CatalogDetail?CSNUMBER=7675">http://www.iso.org/iso/en/CatalogDetailPage.CatalogDetail?CSNUMBER=7675</a> .
ISO ILL	Information and documentation—Open Systems Interconnection—Inter-library Loan Application Service Definition (ISO 10160 and ISO 10161) See <a href="http://www.nlc-bnc.ca/iso/ill/standard.htm">http://www.nlc-bnc.ca/iso/ill/standard.htm</a> .
MAB	<i>Maschinelles Austauschformat für Bibliotheken</i> . See <a href="http://www.ddb.de/professionell/mab.htm">http://www.ddb.de/professionell/mab.htm</a> .
METS	The METS schema is a standard for encoding descriptive, administrative, and structural metadata regarding objects within a digital library, expressed using the XML schema language of the World Wide Web Consortium. See <a href="http://www.loc.gov/standards/mets/">http://www.loc.gov/standards/mets/</a> .
MODS	The Metadata Object Description Schema (MODS) is an XML schema intended to be able to carry selected data from existing MARC21 records as well as to permit the creation of original resource description records. It includes a subset of MARC fields and uses language-based tags rather than numeric ones. See <a href="http://www.loc.gov/standards/mods/">http://www.loc.gov/standards/mods/</a> .
NCC	<i>Nederlandse Centrale Catalogus, NCC</i> . See <a href="http://picarta.pica.nl/DB=2.4/LNG=EN/">http://picarta.pica.nl/DB=2.4/LNG=EN/</a> .
NCIP	NISO Circulation Interchange Protocol (NCIP) is designed to perform the functions necessary to lend items, to provide controlled access to electronic resources and to facilitate cooperative management of these functions. See <a href="http://www.niso.org/committees/committee_at.html">http://www.niso.org/committees/committee_at.html</a> .
OAI harvest protocol	The Open Archives Initiative Protocol for Metadata Harvesting (referred to as the OAI-PMH) provides an application-independent interoperability framework based on metadata harvesting. See <a href="http://www.openarchives.org/OAI/openarchivesprotocol.htm">http://www.openarchives.org/OAI/openarchivesprotocol.htm</a> .

OCLC	OCLC, Inc. is a non-profit membership organization serving 41,000 libraries in 82 countries and territories around the world. See <a href="http://www.oclc.org/about/">http://www.oclc.org/about/</a> .
OCLC PICA	OCLC PICA B.V. European organization of cooperating libraries. See <a href="http://www.oclc-pica.org">http://www.oclc-pica.org</a> .
ONE2	OPAC network in Europe. See <a href="http://www.one-2.org/">http://www.one-2.org/</a> .
ONIX	ONIX is the international standard for representing and communicating book industry product information in electronic form, incorporating the core content. See <a href="http://www.editeur.org/onix.html">http://www.editeur.org/onix.html</a> .
OpenURL	The OpenURL is designed to enable the transfer of the metadata from an information service to a service component that can provide context-sensitive services for the transferred metadata. See <a href="http://www.niso.org/standards/resources/OpenURL-release.html">http://www.niso.org/standards/resources/OpenURL-release.html</a> .
PiCarta	PiCarta is an integrated, multi-material database which contains request facilities and which offers access to online resources and electronic documents. See <a href="http://www.oclc-pica.org/?id=102&amp;ln=uk">http://www.oclc-pica.org/?id=102&amp;ln=uk</a> .
Publiekwijzer	Publiekwijzer is an information service directed at public library users. See <a href="http://www.oclc-pica.org/?id=103&amp;ln=uk">http://www.oclc-pica.org/?id=103&amp;ln=uk</a>
SUDOC	<i>Système Universitaire de Documentation.</i> See <a href="http://www.sudoc.abes.fr">http://www.sudoc.abes.fr</a> .
TEL	The European Library, the gate to Europe's knowledge. See <a href="http://www.europeanlibrary.org/">http://www.europeanlibrary.org/</a> .
UCP profile	The Union Catalog Profile is a protocol over the Z39.50 update service. See <a href="http://www.nla.gov.au/ucp/">http://www.nla.gov.au/ucp/</a> .
UNICODE	Unicode is an entirely new idea in setting up binary codes for text or script characters. Officially called the Unicode Worldwide Character Standard, it is a system for "the interchange, processing, and display of the written texts of the diverse languages of the modern world." It also supports many classical and historical texts in a number of languages. See <a href="http://whatis.techtarget.com/definition/0,,sid9_gci213250,00.html">http://whatis.techtarget.com/definition/0,,sid9_gci213250,00.html</a> .

URL	A URL (Uniform Resource Locator) is the address of a file (resource) accessible on the Internet.
VCUC	The Virtual Canadian Union Catalog. See <a href="http://www.nlc-bnc.ca/resource/vcuc/">http://www.nlc-bnc.ca/resource/vcuc/</a> .
VIAF	VIAF is a joint project with the Library of Congress and <i>Die Deutsche Bibliothek</i> . VIAF explores the virtual combination of the name authority files of both institutions into a single name authority service. See <a href="http://www.oclc.org/research/projects/viaf/index.shtm">http://www.oclc.org/research/projects/viaf/index.shtm</a> .
WorldCat	WorldCat (the OCLC Online Union Catalog) is the world's most comprehensive bibliographic reference resource, with over 53 million bibliographic records representing 400 languages and covering information dating back to the 11th century and holdings information from libraries in 45 countries. See <a href="http://www2.oclc.org/worldcat/">http://www2.oclc.org/worldcat/</a> .
XML query	XML Query aims to provide flexible query facilities to extract data from real and virtual documents on the Web. See <a href="http://www.w3.org/XML/Query">http://www.w3.org/XML/Query</a> .
Z39.50	Z39.50 is an ANSI/NISO standard that specifies a client/server-based protocol for searching and retrieving information from remote databases. It is also an ISO standard ISO 23950. See <a href="http://www.loc.gov/z3950/agency/">http://www.loc.gov/z3950/agency/</a> .
Z39.50 Holdings Schema	See <a href="http://lweb.loc.gov/z3950/agency/defs/holdings-1-0.html">http://lweb.loc.gov/z3950/agency/defs/holdings-1-0.html</a> .
ZDB	<i>Zeitschriftendatenbank</i> . See <a href="http://www.zeitschriftendatenbank.de/">http://www.zeitschriftendatenbank.de/</a> .
ZING SRU	The 'Search/Retrieve URL Service', SRU, is a proof-of-concept initiative to permit the development of value-added search and retrieve applications, such as the scholar's portal, that will integrate access to various networked resources. See <a href="http://lweb.loc.gov/z3950/agency/zing/srw/sru.html">http://lweb.loc.gov/z3950/agency/zing/srw/sru.html</a> .

ZING SRW	The 'Search/Retrieve Web Service', SRW, is a proof-of-concept initiative to permit the development of value-added search and retrieve applications, such as the scholar's portal, that will integrate access to various networked resources. See <a href="http://lcweb.loc.gov/z3950/agency/zing/srw/specifications.html">http://lcweb.loc.gov/z3950/agency/zing/srw/specifications.html</a> .
ZLOT	Z Texas Implementation Component of the Library of Texas. See <a href="http://www.tsl.state.tx.us/lot/ZLOTwhitepaperlib.html">http://www.tsl.state.tx.us/lot/ZLOTwhitepaperlib.html</a> .

# Chapter 2

## The Virtual Union Catalog

Karen Coyle<sup>1</sup>

### 1 Introduction

Some library consortia have chosen to implement a ‘virtual’ union catalog through broadcast searching of the catalogs in their consortium. This is generally a less expensive solution than the creation of an actual union catalog database that must receive and store records from each of the library systems. In most cases it is not possible to do an evaluation of the effectiveness of these two solutions, and therefore a cost-benefit analysis is not available to library administrators who are attempting to make a decision about what type of union catalog best serves their users. Because the University of California had both a centralized union catalog (MELVYL<sup>®</sup>) and a number of contributing systems that were accessible through the Z39.50 search protocol, we were able to do a direct comparison of the retrievals between the union catalog and its ‘virtual’ equivalent. The study showed that the two union catalogs were far from equivalent, and that broadcast searching across disparate databases produces highly inconsistent results.<sup>2</sup>

The University of California is a system of nine (soon to be ten) campuses that span the state of California from Davis, in the north, to San Diego, at the Mexican border—a distance of 800 kilometers. The campuses

---

<sup>1</sup> California Digital Library, <http://www.cdlib.org>, <http://www.kcoyle.net>.

<sup>2</sup> Details of the results of this study were published in D-Lib Magazine in March 2000. See <http://www.dlib.org/march00/coyle/03coyle.html>.

combined have a student enrollment of 160,000, with 10,000 full faculty members and over 130,000 staff and teaching personnel. The campuses function fairly autonomously for most of their academic activities and their administration, although sharing among the libraries is encouraged and well-supported.

The university was founded at Berkeley in 1873 and the Berkeley library is still the largest in the system, with about nine million volumes. The next largest library is Los Angeles, with 7.5 million volumes. The total number of volumes in the 9-campus system is 31 million. There are at least 200 libraries in the system, although this number does not count the many departmental or faculty libraries. Each library has its own unique characteristics. The library at UC San Diego has made agreements with the University of Beijing to receive full-text copies of millions of volumes of its holdings and to make them available to scholars in the United States. The library at Los Angeles has one of the world's largest archives of films, and now serves as an archival agency for some of the top Hollywood studios. Berkeley's rare books room houses the Mark Twain papers; Santa Cruz has an excellent collection of California poetry; Riverside collects contemporary science fiction.

In the mid-1970s the university was seeking ways to make the library's collections more widely available to students and faculty at the various campuses. It was not unusual for a scholar to travel from one campus to another to take advantage of the library collections. The hard part, though, was knowing what you would find there. There was no central catalog for the libraries, so it was necessary to go to the library and consult the card catalog to determine what materials were available. Clearly a union catalog would greatly facilitate the sharing of collections.

Work on a union catalog began in the late 1970s. The first union catalog was a book catalog created from copies of cards contributed by each campus. Before this catalog was completed, a new resource became available: machine-readable records from OCLC, whose card-production service was used by most of the campus libraries. By 1980, the university had produced a microfiche catalog of current cataloging from all nine campuses. But technology was moving forward at a rapid pace, and the key element to delivering machine-readable data directly to the libraries was falling into place: computer networking. The union catalog became a project

of the university libraries that not only created one of the first online catalogs, but also established the first telecommunications network that connected the University of California campuses.

## 2 The MELVYL Union Catalog

I shall begin by reviewing the situation in 1980, when work began on the University of California's union catalog. There were no online catalogs available commercially for libraries; each of the UC libraries maintained a card catalog with cards obtained through the services of OCLC or RLG. Libraries had begun using these services in the mid- to late 1970s, and thus there were machine-readable records for this period only. The libraries did not receive copies of their machine-readable records because they had no use for them. The MELVYL union catalog would therefore serve a dual purpose: it would be a public access catalog for library users, and it would be the archive of machine-readable cataloging for the libraries. Indeed, when the libraries later developed or purchased new systems, those first systems were often created with records exported from the union catalog.

This dual purpose led to a unique design for the union catalog. Where other systems, such as OCLC, kept a single copy of the bibliographic data and added library holdings to this record for additional contributors, the MELVYL developers were obliged to keep all the bibliographic data from the contributing campuses, not just the holdings. Yet they did not want to show a separate record for each campus, since the repetition would be difficult for catalog users. Instead, the design called for a single bibliographic record, with multiple holdings where libraries held copies of the same item. Using an algorithm to determine when incoming records represented the same work, records were then merged into a single record with multiple holdings, but with no loss of bibliographic data. To do this, a composite record was developed based on the USMARC format, but extending it to allow each field to be stored with a digital flag indicating which campus had contributed it.

## Sample record:

- 100 1 Twain, Mark, \$d 1835-1910  
<LC, IG, SDG, LAG, DG, BG, SC, SB, HAST>
- 240 10 How to tell a story. \$f 1996 <BG>
- 245 10 How to tell a story, and other essays / \$c Mark Twain  
foreword, Shelley Fisher Fishkin ; introduction, David  
Bradley ; afterword, Pascal Covici, Jr.  
<LC, IG, SDG, LAG, DG, BG, SC>
- 245 10 How to tell a story, and other essays / \$c Mark Twain ;  
foreword, Shelley Fisher Fishkin ; introduction, David  
Bradley, afterword, Pascal Covici, Jr. <SB>
- 260 New York : \$b Oxford University Press, \$c 1996.  
<LC, IG, SDG, LAG, DG, BG, SC, SB>
- 300 Iix, 233 p., 29 p. : \$b ill. ; \$c 23 cm.  
<IG, SDG, LAG, DG, BG>
- 300 Iix, 233, 19 p. : \$b ill. ; \$c 23 cm. <SC>
- 300 Iix, 233, 29 p. : \$b ill. ; \$c 23 cm. <LC, SB>
- 490 1 The Oxford Mark Twain <LC, IG, SDG, LAG, BG, SC, SB>
- 490 1 Oxford Mark Twain <DG>
- 500 Facsimile reproduction of the first American ed.,  
published New York, Harper & Brothers Publishers, 1897.  
<SDG, LAG, DG, BG>
- 500 Originally published: New York : Harper & Brothers  
Publishers, 1897. <LC, SC>
- 500 Facsimile reproduction of the first American ed.,  
published New York, Harper & Brothers Pub., 1897. <IG>
- 504 Includes bibliographic references.  
<LC, IG, SDG, LAG, DG, BG, SC>
- 504 Includes bibliographic references <SB>
- 505 0 How to tell a story -- In defence of Harriet Shelley --  
Fenimore Cooper's literary offences -- Travelling with a  
reformer -- Private history of the "jumping frog" story --  
Mental telegraphy again -- What Paul Bourget thinks of us  
-- A little note to M. Paul Bourget. <IG, SDG, LAG, DG, SC>
- 650 0 Storytelling <LC, IG, SDG, LAG, DG, BG, SC, SB>
- 700 1 Fishkin, Shelley Fisher <SB>
- 700 1 Bradley, David <SB>
- 700 1 Covici, Pascal <SB>
- 752 United States \$b New York \$d New York \$9 (1996) <BG>
- 800 1 Twain, Mark, \$d 1835-1910 \$t Works. \$f 1996.

```

        <LC, I G, SDG, LAG, DG, BG, SC, SB>
650 0  Storytelling <LC, I G, SDG, LAG, DG, BG, SC, SB>
700 1  Fishkin, Shelley Fisher <SB>
700 1  Bradley, David <SB>
700 1  Covi ci , Pascal <SB>
752    United States $b New York $d New York $9 (1996) <BG>
800 1  Twain, Mark, $d 1835-1910 $t Works. $f 1996.
        <LC, I G, SDG, LAG, DG, BG, SC, SB>

```

One can see from this example that there are multiple versions of many fields with either significant or minor variations (such as the 490 field). There are also fields that were contributed by only one of the libraries, such as the 700 fields contributed only by UC Santa Barbara (<SB>), and the 752 field contributed only by Berkeley (<BG>).

This very complex MARC-like record stayed in the background, and the user of the catalog saw a normal bibliographic display and consolidated holdings:

```

Twain, Mark, 1835-1910.
How to tell a story, and other essays / Mark Twain ;
    foreword, Shelley
Fisher Fishkin ; introduction, David Bradley ; afterword, Pascal
    Covi ci , Jr. New York : Oxford University Press, 1996.
Series title: Twain, Mark, 1835-1910 Works. 1996.
    HAST 5th Stks PS1322 .H6 1996
    UCB Bancroft PS1322 .H6 1996 Mark Twain Papers *c2 copies
    UCB Main PS1322 .H6 1996
    UCD Shields PS1322.H692 1996
    UCI Main Lib PS1322 .H6 1996
    UCLA EngReadRm PS1322 .H6 1996 Main Reading Room ERRREAD-
    STAX
    UCLA YRL PS1322 .H6 1996 Stacks URLSTAX-STAX
    UCSB Main Lib PS1322 .H6 1996
    UCSC McHenry PS1322 .H65 1996
    UCSD SSH PS1322 .H6 1996

```

Only one contributed record was designated the display record; the other records remained hidden from view. However, these other records did contribute to the indexes for the record group. This meant that if one campus had added a unique field, such as the author fields contributed by UC Santa Barbara in the example, a search on that heading brought up the entire group even though other libraries had not included that heading. The merged record became a kind of 'super-record', combining the bibliographic efforts of the whole UC system.

The 'super-record' also had some additional advantages that we had not considered when we were developing the catalog.

The 1980s and early 1990s were given over in many U.S. libraries to the retrospective conversion of their card catalogs to machine-readable form. Libraries were developing online catalogs but only had records dating from their first use of card services like OCLC's. The entire back file of their card catalog had to be converted to MARC records so they could have a complete catalog online. This retrospective conversion was expensive and time-consuming, and in addition was very prone to error. Libraries sent their card catalogs away to be keyed in factory-like settings, and then had to check and correct the records received. Because full-level cataloging for many titles was not available in machine-readable form, some libraries chose to have only minimum-level records created as a way of saving money. This retrospective conversion effort added tens of millions of titles to the OCLC database, however, and collectively the U.S. libraries created the largest storehouse of full cataloging in machine-readable form.

The University of California libraries undertook retrospective conversion at different rates and using different services. Some created mainly full-level records, others were only able to create minimal records for much of their collection. And this is where we discovered a hidden feature of our system's design: as long as one library in the system contributed a record with full cataloging, all others could do a minimal record that would merge with the full one and gain the advantage of the full record in the union catalog. Eventually, most minimal-level records were upgraded by the libraries because they needed full records in their own integrated library systems, but the creation of minimal-level records allowed the libraries to close their card catalogs in a timely fashion and gave them another decade to complete the work of transforming their catalog. At the time of the study reported

here, retrospective conversion was essentially complete and the union catalog held merged records for about 10 million titles, which represented 18 million contributed library records.

### 3 The Virtual Union Catalog

By the early 1990s, each library had its own integrated library system (ILS), and therefore its own online catalog. The systems in place represented three different vendors and a variety of versions among those vendors. These local catalogs fed records directly into the union catalog to create a union copy of the cumulative holdings of the campus library systems. Nearing the year 2000, most of these local catalogs had Z39.50 capability which would allow external systems to send queries to their databases and receive search results. The MELVYL system had developed the capability of broadcasting searches to multiple databases simultaneously and bringing back results for users. So it became logical to ask ourselves: could the union catalog be replaced by a virtual union catalog, that is, a broadcast search across the very same local catalogs that were contributing to the union catalog? It seemed logical to assume that the results of a broadcast search would be the same as the results of a search of the same records in the union catalog. And if that was the case, then a virtual union catalog might be able to replace the current centralized MELVYL database, with a potential cost saving to the University.

We organized a test of this theory. We began with a set of real searches from the logs of union catalog activity. We knew that these searches had retrieved items in the union catalog. We then needed to find out how many records these searches retrieved for each contributing library. Our system allows us to limit our searches by library, so we reran the queries for each library to get the number that we would compare to the retrievals using Z39.50 against their own database.

A retrieval in the union catalog resulting in one record that was a composite of contributions for three libraries would, of course, get no records for the other six libraries, so we needed to create a set of searches for each library that got at least one retrieval in the union catalog.

Although we would have liked to include a wide variety of indexes in our study, it was difficult to find even a small number of indexes that were common among the 6 systems that we would be searching. Many systems had a ‘keyword’ index, but the fields included in this index varied between systems, and MELVYL lacked this field altogether. Some systems allowed only left-to-right searching on certain fields, while others treated those fields as keyword searches. In the end we settled on three indexes:

- Author
- Title (left-to-right search, with truncation)
- Keyword (a combination of title and subject keywords where the system did not have the index).

We then wrote a script that took the searches for each library and sent them as Z39.50 queries to the library’s online catalog. The results were logged for further analysis.

## Results

We fully expected to find some differences in search results based on the unique qualities of the union catalog, in particular the cumulative effect of the merged campus records with the combined retrieval of their headings. In fact, the resulting differences were much greater than we had anticipated, and only a few of them were related to the merging of campus records in the union catalog. Instead, the differences were related to how indexes were structured in the local systems that would make up the virtual union catalog, and the particulars of how searches were performed in the different systems.

To illustrate the flavor of the degree of difference between the search results, consider Table 1, which has positive numbers where the local system returned more records than the union catalog, and negative numbers where the local system returned fewer. Each column represents a library that was queried (L1, L2, etc.):

**Table 1. Author Searches**

<b>Search string</b>	<b>L1</b>	<b>L2</b>	<b>L3</b>	<b>L4</b>	<b>L5</b>	<b>L6</b>
ABBEY	-12	129	-2	-2	-2	4
AURELIUS	307	-155	-211	-213	-197	-313
HAND	462	33	735	1163	868	1973
BRI TTEN, J	-4	-11	-1	-2	-1	-2
BRI TTEN, JAMES	17	-6	0	-1	-1	-1
IMMANUEL KANT	115	-146	-145	-121	-113	-191
LANGSTON HUGHES	19	-91	-64	-64	-86	-103

The searches represent a variety of search types, even though they all are searches on author names. The first three were given just a single name, presumably the family name of the author. The next two are searches in which the family name is given first and is distinguished by the use of the comma; this is followed by a forename or initial. The last two show authors being searched in the form they might appear on a book cover. All of these are legitimate searches on the part of the catalog user.

What caused the differences in search results? After all, these same records are in local catalogs and in the union catalog, so the search results should be nearly identical in both.

Consider the three searches where only a single word was input. In the case of the word ‘aurelius,’ this generally retrieved fewer records in the local catalog than in the union catalog. In the case of the word ‘hand,’ the results were uniformly greater in the local catalog than in the union catalog. Yet both were single word searches against an author index. The explanation for the results in the ‘aurelius’ search is that the union catalog performs a keyword search on author names and therefore ‘aurelius’ retrieves records where that keyword also matches a forename. The local

systems almost uniformly do their searching in a left-to-right manner against a name index that places the family name before forenames:

Aureli us, Marcus

Thus they would not retrieve a record where the author was

Adeodatus, Aureli us

which was retrieved by the union catalog.

The ‘hand’ search is an example of the effects of automatic truncation. In some systems, the author search was automatically truncated so that the search on ‘hand’ became a search on any name starting with the four characters ‘hand.’ So this search would retrieve

Hand, Jacob

Handen, Max

Handers, May

Handschmidt, Frieda

etc. It is not always possible to turn off this truncation in searches and it greatly increases the number of records that any search retrieves.

In the systems that do this truncation it would have also been done for the searches on ‘aurelius’, yet that produced many fewer ‘extra’ results. The reason is obvious: fewer words have ‘aurelius’ as beginning characters than have ‘hand.’ But the difference in retrieval for these two searches is significant, and we can assume that these nuances are not at all understood by users of our catalogs.

Now let us look at some title searches:

**Table 2. Title Searches**

Search String	L1	L2	L3	L4	L5	L6
THE PROCESS	-573	75	289	276	177	392
THE SOCIAL ANIMAL	-7	1	2	0	3	1
THE VISUAL DISPLAY OF QUANTITATIVE INFORMATION	-3	-1	0	0	0	1

Search String	L1	L2	L3	L4	L5	L6
VOI CE	-566	1262	448	497	461	763
WEBSTER' S DI CTI ONARY OF SYNONYMS	-2	0	0	0	0	2

These searches were all done on a left-to-right title index in each system. One source of differences in these results was how the system applied truncation. A system can truncate directly after the last character in the query:

Voi ce#

Or it can add a space and then truncate, creating a word break:

Voi ce #

The first search will retrieve both “Voice of the Master” and “Voices of Our Children.” The second search will retrieve only “Voice of the Master.” Truncating at a word break is often used where truncation is applied automatically by the system after the query is completed by the user. The logic is that few users type in a query that stops in the middle of the last word. What users actually type, of course, has to do with the training they have been given on the system and their experience with results.

Another difference in the title searches resulted from the treatment of articles at the beginning of titles in the indexes, and again at the beginning of queries typed in by users (searches 1–3, above). The MARC21 record considers articles at the beginning of a title to be ‘non-filing’ and these are generally ignored in indexing. So the title “The Magic Mountain” is indexed and filed under “Magic,” not “The.” Users, however, may not always know when to drop these articles in a query. Some of the more clever systems look for the most common of initial articles and remove them from a query if the user includes them. This is imprecise, but it does help some searches which, although they are exact transcriptions of the title, will fail because the user did not know to remove the initial article. In our study, library ‘L1’ clearly was not treating initial articles the way they were treated by the other systems.

We also ran into differences relating to the length of the key that was used for the title index. All systems have some limitation on the length of the title key, but the exact size of the key varies between systems. A longer key means more precision for the user, but it also means more storage for the system. A system with relatively short title keys will retrieve more records for some queries, some of which will be false hits. If the retrieval is not overly large, this merely means that the user must work through some undesired records. If the libraries taking part in the virtual union catalog are large, however, these results could overwhelm the user with unwanted records and make finding the desired records very tedious.

The results of title searching were more consistent with the union catalog, at least in some instances, than the author searching results, and where they differed they tended to retrieve more records in the local library system, whereas the author searches often retrieved fewer. Still, in some instances the differences were significant.

Another source of great differences between systems has to do with what fields have been chosen to populate indexes. Although it may seem that we all know what we mean by 'author' or 'title,' in fact our systems demonstrate that we have taken quite different paths in creating those indexes for our systems.

The MARC21 record that is used in the United States has numerous fields that could be considered titles. There is the title proper, in the MARC 245 field, and there are fields for variations on that title. If a serial, the document may have one or more abbreviated versions of its title. If a monograph, there could be a series title, or two. Any items that have multiple parts, such as a music recording with a variety of pieces, can have titles relating to each of those parts. And other special publications such as conference proceedings have titles for the event as well as the publication. So you can expect that a title index may reflect a wide variety of choices on the part of the librarians who set up that particular system.

The keyword index is equally variable. Some systems index absolutely every possible field in their keyword index, and it is usually understood to be something of a catch-all field, although in America the phrase used is the pejorative 'kitchen sink.' Some system developers may consciously exclude fields that introduce 'noise' but are rarely useful for retrieval, such

as the general notes field. Finding two systems with the same selection of fields in their keyword index would be difficult.

Subject, a search that nearly all systems include, is also problematic. There are the regular subject fields, but there are also fields that have a subject role, at least in the minds of some. The MARC21 record has fields for geographic area covered by the text, for the genre of the item (bibliography, electronic archive, etc.), and additional fields for special collections such as the book binder or the provenance of the item. Are these to be included in the subject index? If not, there may not be another index in which to put them.

Of all fields, it would seem that we share the meaning of the term 'author.' If only that were so. To begin with, there is the question of those authors who are not persons; corporate bodies, institutions and conferences; all have authorship roles that may be recognized by library catalogers, although rarely by the library's users. Users do want to be able to find works using these entry points, although they may not think to search for them in the author index. There are also difficulties defining authors for the less traditional works, such as music or film. Who is the author of a film? Is it everyone whose name is listed in large type in the credits, the producer, the writer, the director? And in the case of a piece of classical music that is performed by a modern orchestra, who or what should be an author-like search point? Composer? Arranger? Conductor? Orchestra? And some systems do not have an author search, but instead a personal name search. This search includes all personal names in the bibliographic record, including those used as subjects. There is a certain logic to this in that a single search retrieves all items by and about a person.

All of these differences contribute to variable results when broadcasting the same search to multiple systems. And this has implications for the creation of a virtual union catalog.

#### 4 Requirements for a Virtual Union Catalog

The scope of this project was not sufficient to provide a full test of functional requirements for a virtual union catalog, but some important

general areas have been identified which would require further analysis and testing prior to planning for the production use of this architecture.

### Database Consistency and Search Accuracy

What our test showed was that the biggest problem in using a virtual union catalog is the inconsistency of results. If all of the systems participating in the union catalog had the same definitions for indexes, did the same normalization of index keys, and performed their searches in the same way, then it would be possible to obtain consistency. This is not the situation in many consortia. It is important, therefore, if you are considering the creation of a virtual union catalog, to study the retrieval capabilities of the library systems that will be included. If you are using Z39.50 to broadcast searches to these systems, you may be able to customize the searches that are sent to each library system to help ensure that the results that you retrieve from the systems are comparable.

This also means that changes to the local systems could affect the union catalog search, so change information must be shared among the library systems.

### System Availability

When you create a virtual union catalog, you are dependent on the system availability of each of the systems in the union catalog. It is ideal to have agreement between the systems that they will be available certain days and hours. This catalog solution creates a great interdependency between the libraries that are participating. If a library is taking down its system for maintenance, it may be necessary to inform other libraries in the system that it will not be available.

### Capacity Planning for Library Systems and Networking

The development of a virtual union catalog design has important implications for local system search capacity and network load. Each search is broadcast to all of the local library catalogs, with the potential that each catalog will then process as many searches as the cumulative total that the libraries

previously handled individually. Network capacity planning would be required to accommodate the increased bidirectional traffic between the libraries.

## Sorting, Merging and Duplicate Removal

Searches issued against the union catalog retrieve a set of records that have been merged to eliminate duplicate bibliographic records, and sorted prior to input into the database. Broadcast searches return a set of records without merging or sorting. Although Version 3.0 of the Z39.50 protocol includes a sort function, few systems currently support this feature. Even with that sort in place, the union catalog interface would have to merge the retrieved sets as well as remove duplicate bibliographic information while maintaining individual holdings data. Because searches across our libraries often retrieve large result sets, sorting and merging is expected to be technologically challenging.

## 5 Conclusion

Do the results of this study mean that a virtual union catalog should not be considered as an option for your library and its partner institutions? Not at all. This study pointed out some of the criteria that must be considered when deciding whether to create a centralized database, as opposed to a virtual union catalog. These can be summarized as:

- The success of virtual union catalogs will increase among libraries with similar local systems and similar cataloging and indexing, and will decrease with differences in those aspects;
- A centralized union catalog may be more costly to create, but it can overcome some of the differences in record quality from different institutions and actually enhance retrieval of minimally cataloged items; and
- A virtual union catalog is highly dependent on the day-to-day functioning of local systems; a centralized catalog needs to receive records from local catalogs but otherwise functions independently.

Each library consortium must decide its goals for a union catalog and weigh this against its budget and technical capabilities. The important thing is to understand the system capabilities and to plan your services around what your system can actually deliver.

# **Chapter 3**

## **The Cathedral and the Bazaar, Revisited: Union Catalogs and Federated WWW Information Services<sup>1</sup>**

Stefan Gradmann

### 1 What This Paper Is Not About ...

In the past 30 years, which have witnessed the advent of library automation, numerous speculations have been published, most of them concerned with either the imminent death of libraries that were seemingly doomed to be replaced by some omnipotent electronic successor, or with “business as usual” proclamations basically stating that libraries—even if electrified to the extreme—would ultimately continue to function the way they had done for centuries.

In the past decade, which has seen the ascent of the Internet, such speculations have been heavily intensified and increasingly focused on aspects of information technology and the information economy exemplified by the information and communication models of the World Wide Web. These speculations have led to sometimes astonishing and radical conclusions and assertions; for example, WWW-based information services such as Google or Yahoo! were supposed to take over library

---

<sup>1</sup> Although the relation of this paper’s title to Eric S. Raymond’s essay on “The Cathedral and The Bazaar” is explained in more detail in chapter 4, it should be made clear from the beginning that the title of this paper alludes to this essay.

functions altogether, or librarians were expected to catalog all quality information on the Internet.

None of these radical changes have actually taken place—and yet, a lot has changed. And the speculative striving to make projections and predictions in this field has certainly been fed by the common feeling that something fundamental is happening to our paradigms and techniques of dealing with information, and to our concepts of information themselves. Still, in a period of profound uncertainty, projections that make use of metaphors of the past to predict the shape of future electronic information landscapes do not, in essence, transcend the intellectual qualities of a Star Trek movie, as tempting as they may be.

The present paper tries to avoid bad library science fiction in general, and predictions as mentioned above in particular. Instead, I assume that we can make hardly any valid statements except those concerning the very near future, but that it may be useful instead to describe as precisely as possible what changes and differing approaches can currently be identified in some fields of scientific information technology and economics, and to try to reach an adequate level of abstraction in the description of such changes and differences.<sup>2</sup>

---

<sup>2</sup> When using the term ‘WWW-based information services’ in this paper, I am referring to services such as the NASA Astrophysics Data System (ADS) or the NEC Research Institute Research Index, as well as to more generic services such as Google or Yahoo. ADS and the NEC Index are well presented and discussed at length in a very detailed presentation given by Gerry McKiernan at the WilsWorld ’02 conference (McKiernan 2002). In the announcement of this presentation on the conference website, the following assertions are made: “In recent years, a number of experimental and operational Web-based information systems and services have emerged that offer advanced and novel features, functionalities, and content. In this presentation, a variety of these innovative services will be profiled, as will their associated technologies. The potential impact of these systems on the development and enhancement of commercial and library information services will also be reviewed and discussed.” However, the latter aspect, although announced, is not really discussed in the presentation itself. The present paper therefore can be seen as a complement to McKiernan’s work, which is very extensive as far as WWW services are concerned but quite restricted as regards libraries. As a consequence, librarian aspects are stressed to a higher degree in the present paper.

## 2 ... And What This Paper Does Attempt

This paper is mainly concerned with the differences between the ways in which information is organized; on the one hand in electronic library catalogs and, more specifically, within electronic union catalogs,<sup>3</sup> and, on the other, in genuine WWW-based information services. The main goal here is to identify some of the fundamental differentiating characteristics, whether in terms of the information entities themselves, the way they are conceptualized or the way they are referenced and their identity is established in their respective contexts, or in terms of the actual modes of collaboration within librarian union catalogs and WWW-based information services.

A better understanding of such differences may in turn help us better understand what actually happens within the overlapping zone between both worlds: whenever a union catalog points to information in the WWW domain, or whenever an Internet search engine encounters catalog applications with their index files and librarian metadata, concepts and mechanisms from two different paradigms of information organization are made to coexist and together create a hybrid setting that can be understood better if the originating contexts of the respective mechanisms are kept in mind. The point here is to identify differences and relevant questions (rather than answers) by describing the often complex relation between electronic union catalogs and WWW-based information repositories, in terms of mutual redundancy, competition, and (sometimes and hopefully) convergence.

---

<sup>3</sup> The term ‘catalog’ is used as a synonym of ‘electronic catalog’ throughout this paper, which is thus implicitly restricted to electronic metadata as part of librarian or WWW-based information infrastructures. The author is aware of the segment of union catalog reality that is thus deliberately excluded from the scope of this paper—on the other hand, a comparison of traditional union card catalogs and WWW-based information services really would not have made much sense.

And if some useful hints can be given at the end of the argument concerning the possible ways for both worlds to coevolve in the near future, this paper will have reached its (modest) objectives.<sup>4</sup>

### 3 The Risks of Pragmatism: 2½ Examples

In order to illustrate the need for conceptual clarification, one that is of practical interest, it may be useful to consider two concrete examples taken from the authors' daily working context. Both examples are concerned with the coexistence of library catalogs and WWW-based information services.

#### “Make the WWW Part of the Catalog”

The first example is concerned with a situation most readers of this paper, at least those from the ‘hybrid’ library world, will be familiar with: the need to present coexisting printed and electronic manifestations of works to library users in a consistent service model, more specifically in the area of printed and electronic journals.

Until recently, holdings of electronic journals have not been systematically integrated into library union catalogs, even though many participating libraries spend increasing sums of money to enable their users to access such resources via licensing agreements. This has led to a situation where libraries have started to build vast link repositories for electronic journals outside their respective OPAC environments and thus, along with these developments, a very impressive repository of electronic journals metadata and of library ‘holdings’ (in terms of license agreements) has been built on a national scale in Germany (e.g. the *Elektronische*

---

<sup>4</sup> It is worth noting that this paper is written from the point of view of a librarian; the author—presently active in the gray area shared by both worlds—has a strong background in the union catalog community, and the present audience are librarians and technicians active in union catalog environments. The paper may thus fail to identify some points that are of specific interest to the W3C community, while it probably overemphasizes issues that may seem completely trivial to those who hold a primarily WWW perspective.

*Zeitschriftenbibliothek* or EZB).<sup>5</sup> From a user perspective, the major unsatisfying aspect of this situation is the fact that, depending on whether a printed or an electronic resource is to be retrieved, different ‘catalog’ environments have to be used. There is no way of retrieving both kinds of resources using one single interface. The problem is common to all ‘hybrid’ library architectures and systematically recurs at all scales—from the context of a single library to the issue of how to relate resources like CORC and WorldCat to each other.

One of the practical responses of the library community to this situation has been to try to integrate as many of the pointers to Internet resources into the library information systems, and thus to make parts of the WWW a part of their catalogs. One of the union catalogs the author of this paper is working with is about to move in that direction. One idea that is currently discussed within this union catalog is to simply add all metadata from EZB (the nationwide repository) to the union database, thus creating holdings data for the participating libraries and thereby ensuring replication of these metadata, together with the ‘holdings’ information, to the participating libraries’ OPAC environments.

However (and quite paradoxically), this creates one specific problem in the case of freely accessible electronic journals such as D-Lib Magazine or First Monday: no license agreements are necessary to access these resources, and as a consequence no library-specific ‘holdings’ information can automatically be generated for these resources. Here again, a practical solution has been devised: simply add ‘holdings’ for all libraries participating in the union catalog in the case of such free electronic resources.

The resulting situation is a practical solution to a specific problem that immediately generates numerous new problems of its own. For example, the use of ‘holdings’ information, which is itself a questionable construct as far as licensed electronic material is concerned, almost completely loses consistency with such an approach. We will come back to this issue as well as to the overall problem of inconsistency later on. At this point it is

---

<sup>5</sup> “Electronic Journals Library” would be a rough English equivalent. EZB can be accessed via <http://rzblx1.uni-regensburg.de/ezeit/>

sufficient to highlight the problematic nature of an approach that tries systematically to integrate pointers to WWW resources in library catalogs.

### “Make the Catalog Part of the WWW”

The alternative (or possibly complementary) approach is often considered when it becomes apparent that library information resources tend to be ignored within the overall information economy of the WWW. The culprit here is the so called ‘hidden Web’; metadata contained in library catalogs are mostly ignored by the leading search engines, for the simple reason that the application layer used to access these records is not transparent for generic WWW technology, and therefore ‘hides’ the resources it should make accessible.

Solutions to this problem are often discussed in terms of making library catalogs more systematically ‘WWW-transparent’ by making catalogs more generally part of the WWW. The overall aim of such strategies is to ensure the presence of metadata from library environments (OPAC or union catalogs) in result sets generated via WWW-based search engines, and to eventually ensure that these sets receive a high ranking because of their high granularity and the quality of the indexing information they include.

While seemingly logical, the consequences of such a strategy could be far from desirable, especially if such an approach were adopted by all major university and research libraries plus a significant number of union catalogs. The first and most striking effect would be extreme redundancy of information, quickly approaching unwanted levels of information entropy; what user would actually want to be overwhelmed by thousands of metadata records pertaining to James Joyce’s “Ulysses” from libraries all over the world when doing a search for “Ulysses” in Google? Moreover, users would then be confronted with result sets that pointed to information objects in very different ways; while in some cases direct access to an information resource via a URL pointer may be possible, in the case of metadata originating from libraries the user would be faced with differing and various types of mediated access, an effect that would certainly put into question the results of a strategy that reveals library resources.

## More Integration Strategies ... and the Need for Distinctions

A third prominent integration strategy deserves mentioning here: the systematic use of library systems as gateways to WWW resources.<sup>6</sup> A more generic, and possibly more appealing, variant of such an approach involves all integration strategies that are built around concepts of open and context-sensitive linking as part of library information infrastructures.<sup>7</sup>

Without going into detail at this stage of the argument, it should be said that any over-pragmatic strategy that simply combines library and WWW resources, yet remains unaware of the fundamental differences of the respective information resources, is unlikely to produce satisfying long-term results. This observation does not question the actual need for integration strategies (and we will come back to this point later in this paper), but rather highlights the extent to which strategies need to be built on clearly established distinctions between the information landscapes we ultimately seek to combine.

The following sections of this paper are concerned with such distinctions. For the sake of clarity I will, in what follows, sometimes deliberately ignore ‘hybrid’ infrastructures. Only after having established the basic, underlying, differences will I reintroduce such hybrid (and mostly secondary) settings.

### 4 Differing Basic Elements and Concepts: Entities, Pointers, Identities

Library and union catalogs on the one hand and WWW-based information resources, such as Yahoo or Google or any repository built on a metadata harvesting protocol [specified, for example, by the Open Archives Initiative

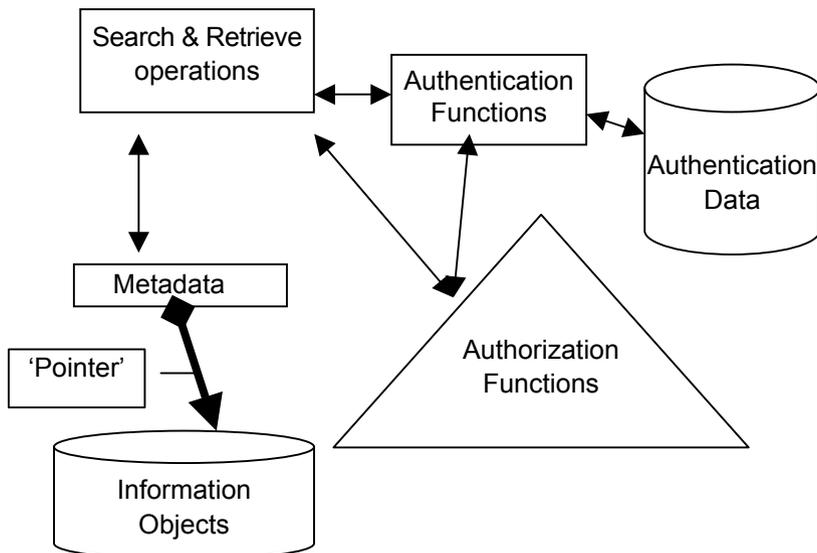
---

<sup>6</sup> S. Thomas has proposed this, for instance, in her reflections on “The Catalog as Portal to the Internet” (Thomas 2000) that have provoked some interesting discussion (cf. Schottlaender 2000).

<sup>7</sup> Such concepts are presented in detail in the contributions from H. van der Sompel mentioned in this paper’s bibliography.

(OAI-PMH)], on the other hand share a number of basic instances and entities as part of their information infrastructure. They mostly contain a distinct metadata layer including pointers to the actual information objects, together with a user interface typically including support for search and retrieval operations. Furthermore, some means of identifying users and information objects must be present somewhere within the respective system: the authentication layer, together with functions that are used to determine what kind of operations a given user (or class of users) may apply to a given information object (or class of objects)—the authorization layer.

From a bird's eye perspective, information systems originating from the library world and from the WWW do indeed have a lot in common. The following diagram visualizes the basic components mentioned above and could be used to describe library information systems and genuine WWW-based systems alike.



However, closer to the ground some basic differences begin to appear. What follows is a closer look at these differences that would be described as ‘distinctive’ (as opposed to variations in detail and granularity).

It may come as a surprise that relatively few of such distinctive/fundamental oppositions can actually be identified in the areas of search retrieval and of ‘bibliographic’ metadata, or that an assumption is being made here that the main differences reside in the ways information objects themselves are conceived, in the way access to these objects is organized and in the mechanisms of authentication and authorization.

However, search interfaces for electronic library catalogs are a relatively young component of libraries and library cooperation, and from the beginning of their short history have evolved much more in line with features and requirements of generic, non-librarian automation technology than, for example, the books themselves, the nature of which has been shaped over centuries long before the birth of electronic information processing.

As for ‘bibliographic’ metadata, the above assumption may be more controversial, especially within the library community; after all, many librarians still regard the production of metadata (in the sense of cataloging) as the very heart of their business, and it may be hard for them to admit that vital issues may well be defined outside the scope of cataloging principles and practice. The assumption is retained nevertheless: many of the guiding principles of cataloging, that had their origins in the sequential organization of card catalogs and that have initially been preserved in electronic cataloging environments, have either vanished or are at least being seriously reconsidered. And even in those cataloging databases that still contain important layers of data oriented towards card catalog production, the creation of a Dublin Core-like interface is comparatively straightforward. This is much easier, anyway, than converting data the other way round; trying to generate traditional cataloging data from a Dublin Core source would probably turn out to be much more of a challenge, if anyone were even interested in the exercise at all.

Furthermore, even the apparently most significant structural differences in the metadata area, such as the ‘holdings’ or ‘copy’ notion of library catalogs that has no real equivalent in WWW-based information services, can be addressed more appropriately as an aspect of pointing and access to

information objects (see below.) And so, while I have devoted a good deal of attention in this paper to the topic of metadata, I will continue to maintain that the crucial differences between Web-based information systems and traditional ones do not lie in this area.<sup>8</sup> Instead, some very evident and fundamental differences can be identified in the remaining three component areas. This involves nothing more than recalling some obvious, though often forgotten, truths regarding the relation of library catalogs and WWW-based information services.

### Books vs. Digital Information Objects: The Basic Information Entities

The first point to be aware of is the profoundly differing nature of the information objects involved. Library catalogs and automation systems are designed to contain descriptive cataloging records for books and book-like printed information, together with pointers to the actual physical copies of these as present on library shelves. WWW-based information systems are designed to contain identifying (and some basic descriptive) information pertaining to electronic information objects (and most typically hyperlinked objects stored somewhere in the network at any location that can be addressed via HTTP), together with pointers to these objects.

It is worth briefly recalling three of the many consequences that have already received their due of scholarly attention. The first consequence is that paper books and other paper publications are combined presentation and storage media, where the display of information is altogether visual and the content is physically tied to the paper and the pages of the publications. On the other hand, with electronic publications storage and presentation are separate. The second consequence is that additional electronic devices are

---

<sup>8</sup> This assumption does not contradict assertions made by the present author in an earlier paper (see Gradmann, 1998). The distinctions made there are less concerned with actual bibliographic metadata than with the respective contexts of use and the originating communities of these metadata.

<sup>9</sup> The contributions contained in TEXT-E 2003 are an excellent starting point for entering the relevant scholarly discussion in the area of both semiotics and information technology.

required for access to the content of digital information objects, whereas books can simply be read using our human senses. Finally, the third consequence is that automated operations on content are possible in electronic information objects in a way that is inconceivable for printed material.

The fact that many digital information objects are still modelled upon the example of printed books should not make us forget the fundamental differences between them: digital information objects will evolve from book-like analogies into new forms of information modeling, forms we do not yet have names for, and this fact is about the only excuse for using such terms as ‘e-books’.<sup>10</sup>

### Shelfmarks vs. Links: The Pointers from Metadata to the Information Objects

The second area where both worlds differ substantially is concerned with the way they organize access to the actual information objects for their respective user communities. To state it simply, library-based information systems are based on the idea of mediated access, whereas the original principle of WWW-based systems is one of direct, instant access. The principal reason for this is the fact that librarian information objects (books and the like) simply are not kept within the information system (the catalog) but on the library’s shelves, whereas in the case of WWW information systems the information objects are technically (or at least can be) part of the system.

This seemingly trivial observation has two very important consequences for the respective architecture of these information systems:

In a library information system, the user is interacting with metadata on all levels: not only with ‘bibliographic’ metadata, but also with a metadata substitute for the real information object within the information system, the copy record, which in turn contains a pointer to some instance outside the

---

<sup>10</sup> For the very same reason, the term ‘digital library’ can be considered as intellectually somewhat dubious: an institution either deals with books (and then can be called a library) or with digital information objects (and why should it then be called a library?).

system that will mediate access to the information object for the user. WWW-based information systems have no equivalent of this ‘copy’ or ‘holdings’ layer, because the information objects themselves are a technical part of the system.

As a consequence, the pointers to the actual information objects have fundamentally different functions within the respective systems: the ‘shelfmark’ or ‘lending number’ pointers point to some instance outside the library catalog (a librarian or a lending module) that will interpret it and finally grant access to the information resource in a way the information system has no knowledge about, whereas the URL pointer (or any technical successor in WWW-based information architectures) basically points to the information object itself that is technically kept within the system (not necessarily stored there physically but part of the system’s technical architecture).

These observations account for numerous functional and technical incompatibilities between library and WWW information systems, and it is important to fully understand their implications before combining working principles from both worlds. The ‘copy’ level of a library system is difficult to translate to the WWW world, and the pointers to the actual information resources react to very different functional requirements.

The latter difference in particular needs to receive additional attention. The ‘shelfmark’ string in the library system may contain almost any information that can be interpreted by humans, from the actual shelfmark (“X 1989/1234” or the like) to information like “go to room 202 and ask there,” or even simply “go and ask the librarian”. And should the copy or call number be erroneous, the lending system module will not recognize it—but ultimately some librarian will be there to help with the matter; the pointer goes outside the system, and the responsibility for resolving the pointing information is outside the system as well. This is the reason why our union catalogs and library OPACs containing such an amazing quantity of incorrect shelfmark information nevertheless continue to function.

The situation is radically different with URL pointers within WWW-based information systems; one character missing in a URL will simply generate code 404 and not reveal any information beyond this error message. Mostly, no external instance can be called upon to correct the pointing information; the correctness and reliability of the pointer are a

vital constituent of the information system. This is why the protocols for constructing and resolving HTTP pointers are relatively strict and elaborate (even though insufficient: there will be successors to URLs as we know them today!) whereas shelfmarks and copy numbers are variable string values with almost no restrictions at all.

Of course, notions of direct access to resources have been added to library-based systems in the recent past, and access control mechanisms and restrictions have been implemented in various ways in WWW-based systems—but still the original governing principles of mediated vs. direct access have been at the origin of the respective systems' design and of the pointing mechanisms used. This is an important fact to remember when one tries to understand what happens to Internet pointers in library systems.

### Identity and Credentials: Authentication and Authorization

Instances that are taken for granted in one information environment may cause near-metaphysical problems in another.<sup>11</sup> This fact can be illustrated with one simple yet striking example (considering the way persons and information objects are identified in both worlds and the way authorization to use a given resource is determined).

In the 'real' world, when trying to establish the identity of a library user, one simple and effective way would be to ask for a passport or ID card. A certain number of additional checks can then be performed; if the ID-document bears the same name the user claims to have and the photograph therein bears at least some resemblance to the owner, and, furthermore, the document has been issued by a trustworthy authority, the librarian may decide that the identity of that user has been established to a sufficient degree. And if that user wanted to borrow a book reserved, for instance, for local residents, a simple check of the address in the user's ID document would quickly solve the issue. Authentication and authorization can thus be established to a sufficient degree using simple and robust techniques.

---

<sup>11</sup> A very sound introduction to the issues of authenticity and integrity is given in Lynch (2000).

However, one of the key factors for the efficiency of this approach is indicated by the words “to a sufficient degree”: the user’s identity is never established with 100% certainty, and there is no need to do so, since a complex set of context information is combined to dynamically evaluate the level of trust required and the degree of certainty needed as a consequence.

The situation gets far more complex once we look at digital authentication scenarios: in this context; identification and authentication information must often be established 100% or simply fails to be established at all. In binary logic, identity is either established or not, and no such notion as “to a sufficient degree” can ease the task. As a consequence it has to be established to a degree that is almost never required in ‘real life’ environments. Or, as Clifford Lynch puts it:

In the digital environment [...] computer code is operationalizing and codifying ideas and principles that, historically, have been fuzzy or subjective, or that have been based on situational legal or social constructs. Authenticity and integrity are two of the key arenas where computational technology connects with philosophy and social constructs. (Lynch, 2000)

And the annoying fact is that this holds not only for persons operating in digital information environments, but for digital information objects as well: the identity and integrity of a printed book is far easier to determine than the identity and integrity of its digital equivalent.

Moreover, while such information is far more difficult to establish in digital environments, ambiguous authentication and identification information can completely block a digital information system, while some flexible strategy of dealing with this lack of information in conventional information environments can almost always be devised.

As a consequence, tremendous efforts have to be made in digital information environments in order to determine what kinds of operations a given user may perform upon a given object, and this places constraints upon the way such environments function, a situation that is almost unknown in ‘conventional’ library contexts.

## 5 Differing Modes of Collaboration: The Cathedral and the Bazaar

In addition to the differences in the two types of information systems mentioned above, important differences can also be located (and must be accounted for) in the way the respective communities cooperate; library union catalogs and federated information environments on the WWW have very different traditions of organizing and experiencing cooperation.

The first striking, and almost trivial, difference concerns the types of cooperating partners: libraries—as different as they may perceive themselves to be—are a more homogeneous group of organizations by far, both in terms of decision making and in terms of user requirements, than the heterogeneous groupings of companies, individual scientists and more or less formally organized parts of the academic community that typically make up the user/production base of federated information services on the WWW.

This basic difference leads to an important secondary observation: rules and guiding principles, as well as common policies for information management, can be imposed much more effectively in a relatively uniform and close user group such as the library sector, while the typical setting within the Internet can never be prescriptive to such a high degree.

The difference is also similar, to some extent, to those described by E. Raymond in his essay on “The Cathedral and The Bazaar” between different modes of collaboration and differing modes of communication when comparing the traditional community of software engineers, for whom the ‘cathedral’-building metaphor is used, and the open source development community, to whom the ‘bazaar’ metaphor is applied.<sup>12</sup> What follows is a brief outline of some of the directions that a closer analysis of this issue should pursue.

If one examined the respective ways in which a WWW development and library staff are collaborating, one would immediately find that the

---

<sup>12</sup> Raymond then goes further than I want to go here: he proclaims the bazaar model to be more powerful than the cathedral model, whereas I have no intention of transposing that conclusion to the context of this paper. This is where the reference to Raymond’s paper has its clear limits.

librarian collaboration model is almost entirely obsessed with rules, whereas such rules hardly play an important role in the WWW environment, where their structural position is taken over by protocols. Likewise, library environments tend to be highly prescriptive as compared to the rather experiment-oriented WWW environments. And finally, library settings seem to have a strong tendency to establish pre-coordinating frameworks, whereas WWW environments tend to assemble collaborative resources first and then post-coordinate their actual collaborative use.

In the field of communication modes, similar observations can be made. Whereas library communities tend towards hierarchical communication models, WWW communities have a rather flat information culture. The channeled vs. broadband perceptions of the communication lines seems to be another relevant distinguishing factor. And one could also argue that the way of organizing communication in libraries is very much oriented towards aggregation of information, whereas the WWW communication paradigm seems to be heavily oriented towards distribution of information, the two worlds thus focusing on two very different aspects of communication practice.

One could even speculate on the differing modes of perception and of mental organization of information units that seem to be at the roots of the respective communities, and might then end up reflecting on the community difference in terms of identity vs. difference, but I will leave such philosophical ruminations for another occasion.

The point now is to create an awareness of the ways in which respective communities differ ‘culturally,’ in their modes of communication and of collaboration. This, together with the insights made in an earlier section, provides sufficient basis for a discussion of possible scenarios for the future relations of these two cultures.

## 6 Modes of Coexistence: Future Choices and Bridging Concepts

### Coexistence? Coexistence!

It should be clear by now how the recognition of the fundamental differences between the two information paradigms helps us to understand better the often unexpected effects produced when transposing objects and

methods from one world to another. While such combinations of objects and methods stemming from very different contexts cannot be avoided altogether and, in order to be sure, must be accounted for systematically in 'hybrid library' settings, it is still useful to keep in mind the side-effects that are produced with such an approach.

The recognition of these differences can also help conceptualise the possible future relation between library catalogs and WWW-based information services, without falling back into the bad habit of excessive and fruitless prediction-making mentioned in the beginning of this paper.

In this attempt to take a modest look ahead, I make two assumptions. First, that libraries with their catalogs and WWW-based information architectures will coexist for quite some time, and even though one paradigm of information organization may eventually gain the upper hand, such a possible future situation is far beyond the scope of this paper. The second assumption is less evident: it is that real choices can actually be made in organizing this coexistence and that the coevolution of both paradigms is not governed by some obscure cybernetic natural law that causes fatal things to happen. The end of this paper is devoted to actual choices we could, and should, make in this area.

### Redundancy, Competition, Convergence, Integration

The possible relations of present and future coexistence can be described using (at least) four different concepts. To begin with, two of these are rather unproductive and ultimately inappropriate. Redundancy may be the least desirable one: modeling the same information objects redundantly in two contexts is expensive, inefficient and carries a high risk of long term overall inconsistency. This is true for all approaches resulting in redundancy, be they based on parallel, unconnected activities in both environments that are not acting in concert in any way, or on data replication scenarios. Competition is not an appropriate characteristic either, even though it may appear inevitably in many political contexts where both paradigms are competing for the same resources (usually money) and therefore are perceived as functionally and technically competing, although they serve fundamentally different needs. Two other characteristics could be more fruitful and may help to establish productive and realistic objectives. Provided the fundamental conceptual

differences between both paradigms are well understood, their relation could evolve either in terms of convergence or of integration. Convergence in this context would mean that both worlds move towards the same objectives, getting continuously closer to each other and possibly creating more and more overlapping areas without, however, blending both paradigms altogether. Catalogs and WWW-based information systems remain clearly discernible worlds in this approach. Integration, in contrast, would mean that both worlds are actually blended into something new, embracing both paradigms and serving the needs of their respective communities in one common approach of information modeling.

Examples of all four characteristics on organizing coexistence can be found in our present professional experience, and most readers of this paper will be able to quickly identify examples of redundant, competing, converging, or integrating scenarios in their own working context. The author of this paper is convinced that (at least) these four scenarios of coexistence will remain valid options in the short term, and that it is up to the stakeholders of both worlds to make their choices among them. Such choices will be triggered by many factors: money, politics, economic interest, to name just a few powerful ones outside the scope of what readers of this paper will typically be able to influence. There are, however, two concepts in the area of information architecture that may help to orient this coevolution in the direction of convergence or integration, and the promotion of these two concepts would be a very useful contribution of the union catalog community to the shaping of future cooperative scenarios.

### Bridging Concepts: FRBR and the Semantic Web

Two important bridging concepts in that sense might well be the metadata layering model expressed in IFLA's "Functional Requirements of Bibliographic Records" (FRBR) and concepts currently taking shape in the "Semantic Web" approach.<sup>13</sup> The general reason is that both concepts raise the level of abstraction concerning information entities that are present in

---

<sup>13</sup> This assumption is by no means meant to be exhaustive: there are certainly more examples of bridging concepts, and the author merely tried to identify two prominent ones.

both information paradigms sufficiently high in order to potentially embrace both worlds, and thus may play an effective bridging role.

Semantic Web technology and, more specifically, methods based on semantic Web ontologies are likely to make new and productive use of the fine-grained semantic metadata that libraries have been traditionally producing, thus enhancing the taxonomies of semantic Web ontologies. Assertions based on the use of classifications and indexing schemes could easily be transposed into taxonomic elements that, in turn, greatly broaden the basis to which inference rules can be applied. This results in a much richer taxonomic base for ontological operations, and could well generate an ongoing process of library work being fed into semantic Web ontologies.

Likewise, the integration of semantic Web techniques in library catalogs, not only for search and retrieval operations but also, for instance, to generate proposals for classification attributes using inference rules, may well help a lot in everyday library work: a rule of the type “If a work by a given author has a given classification element associated with it and if the publication year of another work by an author with the same name is adjacent, the same classification element is likely to apply to this item” would probably yield useful and time-saving classification proposals for newly cataloged items.

It is assumed here that semantic Web-based approaches will primarily contribute to the dynamics of convergence.

The FRBR model that results in a layered metadata architecture has the strategically important advantage of making possible a combination of metadata architectures typical of library union catalogs (and as discussed above in section 3) and of the ‘flat’ metadata models that are typical for WWW information architectures. As a consequence, applying FRBR-based approaches to the development of their catalogs, librarians could substantially decrease the annoying effects that were described above and that today contribute to keeping library metadata resources within the ‘hidden Web’.

Establishing coherent unified concepts of what semantic entities, expressions/manifestations and item derivatives actually are and relating these in one model that makes ‘hybrid’ information settings appropriately conceivable is one of the major advantages of FRBR. Clearly, approaches based on the FRBR model probably have a very high integrative potential.

To conclude, while it does not seem very wise to predict future developments too emphatically, library and WWW communities would probably be well advised to invest concerted efforts in semantic Web technology and in hybrid information models based on the FRBR-approach.

## References

- Gradmann, Stefan. "Cataloging vs. Metadata: Old Wine in New Bottles?" In *64th IFLA General Conference August 16–August 21, 1998, Proceedings*. Online at <http://www.ifla.org/IV/ifla64/007-126e.htm>. Also in *International Cataloging and Bibliographic Control* 28.4.1998: 88–90.
- Lynch, Clifford A. "Authenticity and Integrity in the Digital Environment: An Exploratory Analysis of the Central Role of Trust." *Authenticity in a Digital Environment*. Washington: Council on Library and Information Resources, 2000: 32-50. Online at <http://www.clir.org/pubs/reports/pub92/lynch.html>.
- McKiernan, Gerry. "eProfiles. Innovative Information Systems and Services." Online at <http://www.wils.wisc.edu/events/wworld02/present/eProfiles.ppt>.
- Raymond, Eric S. *The Cathedral & the Bazaar*. Beijing [etc.]: O'Reilly, 1999.
- Schottlaender, Brian E. C. *Commentary on "The Catalog as Portal to the Internet" by Sarah E. Thomas*. Washington: Library of Congress, 2000. Online at [http://www.loc.gov/catdir/bibcontrol/schottlaender\\_paper.html](http://www.loc.gov/catdir/bibcontrol/schottlaender_paper.html).
- Text-e: *Le texte à l'heure de l'Internet*. Sous la direction de Gloria Origgi et Noga Arikha. Paris. BPI, 2003. Online at: <http://www.text-e.org/>.
- Thomas, Sarah E. "The Catalog as Portal to the Internet." Washington: Library of Congress, 2000. Online at [http://lcweb.loc.gov/catdir/bibcontrol/thomas\\_paper.html](http://lcweb.loc.gov/catdir/bibcontrol/thomas_paper.html).

- Van de Sompel, Herbert, and Oren Beit-Arie. "Open Linking in the Scholarly Information Environment Using the OpenURL Framework." *D-Lib Magazine* 7:3 (March 2001). Online at <http://www.dlib.org/dlib/march01/vandesompel/03vandesompel.html>.
- Van de Sompel, Herbert, and Oren Beit-Arie, "Generalizing the OpenURL Framework beyond References to Scholarly Works: the Bison-Futé Model." *D-Lib Magazine* 7:7/8 (July/August 2001). Online at <http://www.dlib.org/dlib/july01/vandesompel/07vandesompel.html>.



# Chapter 4

## Linking in Union Catalogs

Ole Husby

Identifying and categorizing relations is a necessary requirement for the formal description that makes navigation possible in the bibliographic universe.

– Knut Hegna

### 1 Introduction

Lately, we have seen a number of new developments of union catalogs, regarding both their content and technical implementations. New material types are making their way into these catalogs/databases, the most notable perhaps being the network documents, residing somewhere on the Internet instead of on a library shelf. It naturally follows that new ‘document delivery’ mechanisms are in demand, and that the notion of holdings needs to be redefined.

In addition, so-called digital libraries are emerging, more or less to complement or even to include the services offered by the library catalogs. A core element of these digital libraries is a technology for managing, expressing and navigating relationships. However, the same importance should be attributed to relationships in library catalogs and union catalogs, especially when available on the Web. This paper will discuss a number of aspects of bibliographic relations and their expression, here called linking. For the purpose of this paper, I will take the liberty of using the term ‘union catalog’ in the broadest possible meaning, not even trying to offer a definition.

Without jumping to conclusions, I will claim that there is a trend within many information services towards the use of more dynamic data models and technical solutions, allowing relationships to be investigated or synthesized as the information space emerges or is explored. A popular description of this is ‘just-in-time’ links, as opposed to ‘just-in-case.’

A lack of linking facilities might lead to replication of data and cataloging effort, redundancies and inconsistencies in the data structures, and more cumbersome tasks for users to collect together items that belong together and to distinguish between items that do not. The basic need for this collecting and separating was especially needed in the card catalog, but is still of vital importance, at least in all the cases where we are burdened by the much-discussed information overload.

Linking allows iterative information seeking, where the selection of specific manifestations or the selection of the desired (appropriate) items should be separated from, and appear at a later stage than, the topical discovery processes.

## 2 Search Portals, Cross-Searching, Union Catalogs

Since numerous different search portals are being offered, we also have to consider whether the service expected from a union catalog is best obtained by the distributed search paradigm, as we are doing when using a standard network protocol like Z39.50 to synthesize virtual union catalogs. The other option is to stick to the ‘real’ union catalog in the physical sense, but now perhaps with the possibility of using other record collecting mechanisms, such as the harvesting protocol offered by the Open Archives Initiative (the OAI-PMH protocol).

While choosing models and technologies for cross-searching and unifying services is not the topic of this chapter, let me just point out that the needs for such efforts will not disappear, as users increasingly expect to find everything they need at the ‘one-stop shop.’ But there is certainly no single solution that should be recommended for all purposes.

Another apparent development is that the criteria for including diverse document categories in the union catalog need rethinking: “In which database does this record belong?” For electronic journals one might prefer

to have a separate e-journal database, for freely available network documents the subject portals could be suitable alternatives, and so on. This question will not be further treated here, but there is no single answer to this question either.

However, what in my opinion is important in this context is that the need for linking is apparently greater than ever before.

### 3 What is a Link?

Several definitions of links are available, among others:

- A link is an expression of a relation;
- A link is a connection from one page to another destination, such as another page, or a different location on the same page; and
- A link is underlined and blue.

As the latter two are, in my opinion, too strongly tied up with the World Wide Web way of thinking, I will choose the first one. With slightly different wording, we could say that a link is an instantiation of a relation—in the hypertext language, we could call it traversable or executable.

This definition next requires some discussion of relations. In general, a relation represents a meaningful connection between two or more entities. The concept of a relation can be rigorously defined in mathematical terms, which, however, I will not do, since in any case there does not seem to be too much variation in how we understand this term. Neither will I go into the typology of relations, but briefly mention that there are several ways of classifying them. One interesting aspect is to distinguish between

- relations that are a priori given by the nature of things;
- relations that are made up by us, and
- relations that are deduced from statistics.

## 4 A Link is an Expression of a Relation

As mentioned above, this is my preferred definition of a link. There are many ways of expressing it, and not all are hypertext links! Here are some quite different methods:

- Citing together;
- Explicitly stating in text (“See”);
- Using controlled vocabularies;
- Data modeling (relational databases);
- Sharing metadata (identifiers, etc.);
- Linking in hypertext systems (blue, underlined ...).

In traditional thinking, linking has been seen as manifestations of relations between bibliographic records. This brings us next to the catalog.

## 5 Linking in Library Catalogs

In library catalogs, information about books, journals and other information entities are made available to the public. Use of these systems, however, often demonstrates that it is not as easy and intuitive to locate the relevant information as we would like to think. A well-known problem is the failure of these databases to bring together objects that belong together, like translations of a given document into different languages, or the representation of the document in different media. There is a need for a conceptual model that captures the entities and relationships of primary concern to information retrieval.

Nearly all current catalogs describe (by the use of document surrogates) manifestations. This does not imply that the manifestations are the only entities present in the catalog, but rather that the descriptions of other more abstract entities are distributed in a different way. Multiple-item entities can be listed in one record, and information related to the expression and work entities will often be replicated in multiple records. One example is the MARC uniform title element, which acts as a sort of work title in the FRBR (“Functional Requirements for Bibliographic Records”) model. However, this ‘work’ is not present as a distinct and identifiable entity in the catalog.

Maps for describing the entities and their interrelations may be constructed and integrated with the catalog in different ways, either as tightly integrated static parts of the records in the database, or else as a distinct and separate information service to be applied dynamically at runtime, as a separate link service. I will return to this topic later on.

## 6 The FRBR Model

The FRBR model has already been mentioned. And while most library catalogs are as yet quite unmarked by this important effort, it is evidently having a major impact on how system designers are modeling the next generation of bibliographic databases. An indication of the model's success is that it has been warmly welcomed far outside the community of theoretically inclined catalogers.

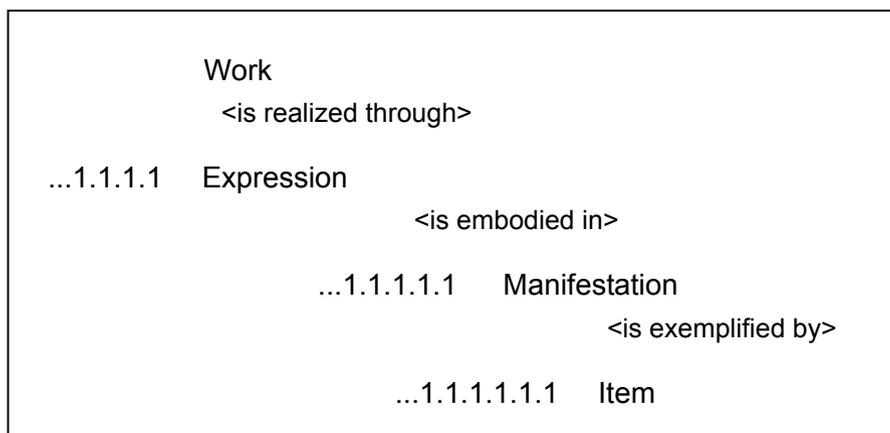


Figure 1. The Product Entities of the FRBR Model

“Functional Requirements for Bibliographic Records” (FRBR) is the title of a report from an IFLA study group, published in 1998. Briefly, FRBR presents a model for bibliographic data based on the entity-relationship data model. Three different groups of entities are defined:

- Group 1, the products of intellectual or artistic endeavor that are named or described in bibliographic records: work, expression, manifestation, and item;
- Group 2, entities responsible for the intellectual or artistic content, the physical production and dissemination, or the custodianship of such products: person and corporate body; and
- Group 3, entities that serve as the subjects of intellectual or artistic endeavour: concept, object, event, and place.

As an example, a work can be a novel, identified by a uniform title. Translations into different languages give the expressions, identified by title. A certain expression can be published on different media, giving the manifestations. Manifestations are often identified by ISBNs. The separate copies are then the items, identified (in the library space) by library codes and shelfmarks.

## 7 The FRBR Relations

The FRBR model further considers the following categories of relations, where some quite simple examples are given:

Between Work, Expression, Manifestation, and Item:

E2 <is translation of> E1

M1 <is manifestation of> E1

To Persons and Corporate Bodies:

P1 <is author of> W1

I1 <is owned by> C1

To Concept, Object, Event, Place:

W1 <is about> C1

Between Persons and Corporate Bodies:

P1 <often cites> P2

P3 <is often cited together with> P4

Between Concepts:

C1 <i s subspecies of> C2

## 8 FRBRizing the Catalog

There is, in my opinion, potential for substantial improvement in library catalog linking, depending on the success of implementing the FRBR model in current and future library catalogs. So what can we do in practice? Here are some directions that should be investigated further:

- Rearrange bibliographic records (entities) according to FRBR, by changing the physical data model, or by extracting FRBR entities at runtime (just-in-time FRBR). This could mean having different types of records, corresponding to the product entity types in FRBR;
- Develop a sound framework for linking, if possible by using externally maintained link services or link databases;
- Choose record identifiers that support linking with as global a scope as possible. There are a number of standard identifier schemes for manifestation records, but hardly any for expressions or works. If we are concerned about interoperability and resource sharing, perhaps we should get together and invent new ones?
- Build maps, paths and navigational tools that guide the user in this new terrain. We have to take into account that user requirements and preferences differ strongly. Nowadays, many of us are used to having result sets sorted and arranged by search engines—often by non-intuitive and ‘magic’ clustering and ranking algorithms. We do not want magic systems (do we?), but comprehensible and accountable system behaviour.

It is not proven that the FRBR model is perfect, and the ideas above are far from easily realized. Experiments have shown that the automatic extraction of FRBR entities and relations may be a very tough task. But the time has not yet come to give in!

## 9 New Opportunities (and Needs)

Whichever definition of a link we might agree on, linking has become a new way of thinking that has emerged with the Web and hypertext. And whatever I might claim, links are perceived as “underlined and blue”... The new opportunities are welcomed and taken for granted by the users. It's now ‘up to the user to click.’ The omnipresence of the Web has already raised users’ expectations with respect to linking everything together: OPACS, A&I databases, e-journals and other full-text archives—even the whole Internet.

Links are sometimes treated as entities themselves, especially within the digital libraries. Separate link databases are flourishing, like SilverLinker, CrossRef and other commercially available services, together with a diversity of proprietary solutions. It should be noted that most of these are ‘closed’ or ‘static’ in some respect.

## 10 Web Services

The Web today (still) depends on us to use browsers to access information services, then to manually parse and analyze the displayed data, in order to identify the roads leading to the desired goals.

Now Web services are here. Web services are strictly organized and standardized Web applications that can be used by other network applications (not browsers) in order to perform a limited task. (These are commonly called business-to-business applications, as opposed to business-to-user.) In general, the use of Web services ought to offer us new ways of integrating and tailoring our information systems, better modularization and extended possibilities for the reuse of tools and services. A potentially interesting area for the application of Web services should be within linking, where separate link services can be accessed by other applications such as the library catalogs on the Web.

## 11 Reference Linking

Today, there is a huge demand for user-friendly methods for reference linking. This is the class of links that can be somewhat vaguely described as linking from metadata (reference, citation) to the full content.

The source may be a metadata record in a database or citation (more or less formally expressed) within some document. The target (full content) may be ‘anything, anywhere’ with a network identifier.

Some common examples of reference links:

- From an A&I database record to the full text;
- From a citation included in a document to the full text;
- From an OPAC record to an e-journal TOC with further linking possibilities;
- From an OPAC record to a network full-text manifestation of a document.

## 12 Static vs Dynamic Links

Most linking architectures are static, in the sense that the links are precomputed (‘just in case,’ ‘a priori’), the target space is a controlled environment, and the links are more or less ‘foolproof.’ On the other hand, we might describe dynamic links in this way: These links are created ‘a posteriori’ (just in time), the target space need not be controlled, and the links are probabilistic (they might not work). Certainly dynamic link creation can include link verification, but this probably takes too long in most applications. (And it seems that automatic link verification can never be 100% reliable.)

As a real-world example of a static reference linking service, I can mention CrossRef:

- This linking service is operated by PILA (Publisher International Linking Association);
- CrossRef is implemented as a static link database;
- The link targets are DOIs (Digital Object Identifiers); and

- Access to the DOI resolution database (metadata → DOI) requires PILA membership.

### 13 Extended Service Links

Reference links usually target one specific copy of the full-content entity. But the user might rather need or prefer

- full content from another supplier;
- an OPAC holdings description;
- a copy ordering/ILL service;
- another metadata description/abstract;
- a book review or access to a net bookshop;
- a ‘full Web’ search.

These are often described as extended services.

It is obvious that every conceivable link is not appropriate to the user, because of

- diverse personal preferences (formats, delivery options, etc.);
- diverse institutional preferences;
- access restrictions, and/or
- temporary unavailability.

These and other parameters constitute the context of the user. The appropriateness of the link depends on this context.

### 14 Closed vs Open Linking

By closed links we understand links that are not context-sensitive:

- They might not work (because of access restrictions);
- They ignore the policy of the user's library; and
- They ignore the user's ‘real’ needs and preferences:

By contrast, we will use the term ‘open links’ for those that are context-sensitive. And furthermore, this means that open linking architectures support extended services

One well-known and pioneering implementation of such an open linking architecture is SFX (‘Special Effects’), which is now a part of the MetaLib product from Ex Libris.

## 15 OpenURL and Service Components

OpenURL is often considered to be a framework for implementing open linking. But strictly speaking, the OpenURL itself is just a standardized syntax for encoding metadata for a document into a URL “... to enable the transfer of the metadata from the information service to a service component that can provide context-sensitive services for the transferred metadata.” There may be several service components available, offered by different agencies or service suppliers. The OpenURL is presently under consideration as a NISO standard.

An OpenURL may look like this:

```
http://www.bi.bsys.no/ourl?si d=BI BSYS: ERL&i ssn=1234-5678&date=1998&volume=12&i ssue=2&spage=134
```

The different parts of this URL must comply with the syntactical requirements of the OpenURL standard.

## 16 Open Linking in Library Catalogs

The preceding discussion may appear to be targeted mostly towards digital libraries. But even in library catalogs, we can consider

- the use of open linking architectures;
- support of OpenURL;
- implementing separate link services;
- support of Web services.

We are seeing the first attempts to implement OpenURLs and separate link services in library catalogs. Web services will surely follow. It is my

hope that such measures may lead to more user-friendly and interoperable systems.

## References

- Functional Requirements for Bibliographic Records: Final Report*. UBCIM Publications, New Series, vol. 19. Munchen: K.G. Saur, 1998. Available at <http://www.ifla.org/VII/s13/frbr/frbr.pdf>.
- Gardner, Tracy. "An Introduction to Web Services." *Ariadne* 29 (2001). Available at <http://www.ariadne.ac.uk/issue29/gardner/>.
- Hickey, Thomas B., Edward T. O'Neill, and Jenny Toves. "Experiments with the IFLA Functional Requirements for Bibliographic Records (FRBR)." *D-Lib Magazine* 8(9). Available at <http://www.dlib.org/dlib/september02/hickey/09hickey.html>.
- Tillett, Barbara B. "Bibliographic Relationships". In *Relationships in the Organization of Knowledge*. Carol A. Bean and Rebecca Green, eds. Norwell, MA: Kluwer: 19–35.
- Van de Sompel, Herbert, and Oren Beit-Arie. "Open Linking in the Scholarly Information Environment Using the OpenURL Framework." *D-Lib Magazine* 7 (3). Available at <http://www.dlib.org/dlib/march01/vandesompel/03vandesompel.html>.
- Aalberg, Trond. "Navigating in Bibliographic Catalogs." In *Proceedings of the 6th European Conference on Research and Advanced Technology for Digital Libraries (ECDL 2002)*. Springer Lecture Notes in Computer Science, vol. 2458: 238–250.

# Chapter 5

## Linda: The Union Catalog for Finnish Academic and Research Libraries

Annu Jauhiainen

All Finnish academic libraries and a number of other research libraries have had the advantage of using a joint library management system for over a decade. A unified network called Linnea was created in the early 1990s, consisting of local installations and a common physical union catalog which were all connected by the powerful and reliable academic data transmission network FUNET. A new library system, Voyager, which replaced the VTLS system in 2001, added new features to the union catalog and makes both cataloging and localization easier and faster.

### 1 History

The academic libraries in Finland have a long history of cooperation in the field of cataloging and library automation. The basic policy has been to follow standards and adopt a joint approach. Since 1977, the libraries have used the FINMARC format and the LSP application purchased from the British Library for offline cataloging and production of printed and microform catalogs. Online databases were already built from these data in the early 1980s. The first union catalogs, one for serials and the other one for foreign monographs, were already created at that time.<sup>1</sup>

---

<sup>1</sup> Esko Häkli, *Off the Record 2. Articles and Papers* (Helsinki: Helsinki University Library, 2000: 98–99).

A new era started in the 1980s. The Ministry of Education funded a project to select a joint automated library system for all academic libraries. The selection process was handled by the Automation Unit of Finnish Research Libraries, a unit within the Ministry established in 1974, which was also responsible for LSP usage on behalf of the libraries. The contract with VTLS Inc was signed in April 1988.<sup>2</sup>

In 1993 the Automation unit with all its tasks and resources was moved to the National Library, where the Division of Database Services continues its work. The unit manages the Linnea network, functioning as a common agency for the academic libraries. In this capacity the National Library is also responsible for the new steps toward Linnea2, as the next generation network is called. The Division of Database Services also maintains the national and union catalog databases.

At the turn of the decade and in the early 1990s, VTLS was implemented in the library databases, one by one. In 1993, when all library databases were up and running, the next step in the Linnea network was to create an online union catalog using the VTLS software. Different options were evaluated. Some people strongly pushed the virtual union catalog option, for they saw the physical union catalog as a waste of money, the money that they would rather have used for local needs. The decision, however, was to go for the physical union catalog, which would be updated by and linked to the local databases. Due to the large number of databases and relative slowness of the FUNET network at that time, a virtual union catalog was not a feasible option. Another reason for establishing a physical union catalog was that the HP3000 servers hosted by libraries would have been heavily overloaded by additional queries generated by a virtual union catalog.

Before the data could be loaded into a union catalog, some customized software development was needed for the VTLS system. For example, a duplicate control algorithm was designed in Finland and subsequently implemented by VTLS. This code was later used in other VTLS-driven union catalog projects, e.g. in Spain and Poland. VTLS also developed

---

<sup>2</sup> Esko Häkli, "A Unified Automation System Using VTLS for Academic Libraries in Finland," *Program* 26/3 (July 1992): 239–248.

features that enabled the libraries to use the Linda union catalog database efficiently for copy cataloging purposes.

The cataloging process was as follows:

1. The record was first searched in the union catalog Linda.
2. If it was found in Linda, it was copied to the local database by entering a single command. In the local database it was possible to do some further editing, e.g. certain fields could be added to the record, etc.
3. If the record was not found in Linda, it was first cataloged there. From there it was copied to the local database, where it could be edited further.

Depending on the material, 50–90% of MARC records could be copied from Linda. Inter-library loan (ILL) localization was also very efficient, because Linda contains summary-level serials holdings from over 400 Finnish libraries. But in the old Linda there were no links between the union catalog and the local database for retrieval of up-to-date holdings and item information. The technology of the time did not make that possible. When you searched a title in Linda, you got the bibliographic record plus a list of libraries holding that title. There was no way of seeing how many copies the libraries had and what the status of the copies was. It was necessary to log onto the local database in order to see the status. The link between Linda and the local databases, which permitted easy copying, was available only in cataloging.

## 2 Linda and the Other Linnea Union Catalogs

The Linda database is the union catalog for the Finnish academic libraries. The numerous libraries of the 20 universities in Finland, along with the Library of Parliament, the National Repository Library and some special libraries, contribute their records to the database. The National Bibliography Fennica, complete from 1488 onwards, is also included in Linda. In addition Linda contains summary-level serials holdings from hundreds of special libraries and polytechnic colleges. Altogether there are over 460 libraries contributing their records to Linda in one way or another.

At the end of 2001, Linda contained 3.7 million bibliographic records, the annual growth being about 200,000 records. The database includes references

on monographs, serials with summary holdings information, cartographic materials, audiovisual materials, electronic resources, multimedia and archives.

Linda does not cover music materials. They are cataloged in Viola, which is the Finnish National Discography and National Bibliography of Sheet Music as well as the union catalog of music materials. Viola contains references to Finnish sheet music since 1977, and to sound recordings since 1901, that is, from the very beginning. Cataloging covers the whole sound recordings and scores as well as the individual compositions contained in them.

In addition to Viola, Linda has another sister database, Manda. Manda is the union catalog of 20 regional central public libraries in Finland. Manda contains references on books, music, visual materials, cartographic materials etc., but not serials, as information on serials holdings of these libraries, as well as many other public libraries, can be found in Linda.

### 3 Selection of the New Library System

Towards the latter part of the 1990s, it became evident that the VTLS system had to be replaced. VTLS had been a trustworthy companion of the libraries for a long time. The system had been a good and stable housekeeping tool, taking care of most of the traditional activities and functions of the libraries. However, that was no longer enough. Due to the rapid change of technology and the new needs in the library field, a new library system was needed, one that could respond to these new requirements and go beyond being a mere housekeeping tool. In answer to the demands of the market, all library system vendors were developing so-called third-generation library systems with relational databases and client/server technology, graphical user interface and Web gateways, the ability to search multiple databases simultaneously, multimedia support and support for internationally accepted standards such as Z39.50, Unicode and ISO ILL, to meet the growing needs of the users. It was also clear that the classic VTLS system was coming to the end of its life-cycle and would not be developed further, since VTLS, Inc. was concentrating on its new system, Virtua.

Because of the great success of Linnea1, as the old network is now called, there was no need to revise the basic service philosophy when moving to a new system. It was self-evident that we would continue with a joint system. Libraries were satisfied with the system and the workflows and cooperation with one another. However, the libraries were open to totally new technical and organizational solutions if these proved more advantageous both functionally and economically. And the National Library wanted to avoid transplanting old patterns into a totally new environment, and wanted to make full use of the advantages offered by the new technology. Thus, we wanted to explore different options for the future database or network architecture during the software evaluation process. One of the important issues was whether to merge existing databases or to keep the current structure. In the RFP, the vendors had been asked if their system could support other kinds of database solutions, i.e. a single central system with full functionality and no local systems, or a data warehouse-type central system of bibliographic data with local circulation systems and indexes. This was also discussed in detail in the negotiations with the final candidates.

Merging existing databases together was technically possible. Some of the vendors even encouraged it. In some cases, it would have meant a significant saving of money in the software price, as well as in the ongoing maintenance of the software and hardware and the overall maintenance of the system. On the other hand, it would have meant a difficult and time-consuming implementation, plus higher implementation costs. Most importantly, it would have meant losing all the work that had been done in the Linda database over the years, because the new centralized database would have to be created from the local databases, not Linda, since the necessary holdings and items information did not exist in Linda but only in the local databases. Besides, we were not convinced of the functionality or the technical merits of such an action, nor the security of the results. We also had to take into consideration the opinion of the participating libraries, which were quite reluctant to pass the maintenance and configuration of their database to a centralized agency that would not be so well acquainted with local customs and needs.

The conclusion of the discussions and the research in this area was that we would gain nothing by merging databases into one centralized system.

On the contrary, it would have made life more complicated, and thus it was decided to keep the same number of databases as before. The same result had been envisioned in the future scenario of the Linnea network that had been prepared at the Helsinki University Library in 1997. According to the scenario, the future network would be based on the Z39.50 and ISO ILL protocols. Use of these standards would allow patrons and staff to log on to different library systems, search remote databases in Finland and abroad seamlessly and retrieve records from them online. This would give new scope for the architecture of the network. According to the scenario, it is likely that the three bases of the network, the local services, the central system and the network connecting them (the Finnish Universities and Research Network FUNET) would remain the same, or almost the same, for the next few years.<sup>3</sup>

The future of the Manda database was reviewed during the selection process. The question was whether to migrate Manda to the new system as an independent database or to merge it into Linda. We also considered freezing Manda as it was, in which case new records would subsequently be added to Linda. As the result of research among Manda users and the feedback from the public library sector, it was decided to continue with Manda as an independent database and migrate it to the new system. We have to admit that we were worried about the quality of Manda records and about what the effect of such merger would be on the quality of Linda, since the Manda libraries use various management systems. These systems have not even always used the MARC format for cataloging, which has not been as standardized in that sector as in the academic libraries.

Due to the obvious benefits of the existing physical union catalog, the issue of a virtual union catalog versus a physical union catalog was not seriously considered. After abandoning the centralized database option, it was a natural choice to continue with the physical union catalog, but with the help of various virtual union catalogs, e.g. subject-based, regional, union catalog of union catalogs (Finnish national databases as well as the Scandinavian Virtual Union Catalog, SVUC).

---

<sup>3</sup> Information Technology Scenario for Finnish Libraries, 1997–2006. See <http://www.lib.helsinki.fi/skenaario/skenarioe.html>.

FINMARC has been the cataloging format of the academic libraries since the 1970s. It has been the basis for cooperation in cataloging. Now that the library system had changed, it was a good time to review the format issue and decide whether to continue with a national format, or to harmonize and go towards a global solution. We saw the advantages of a global option in copy cataloging and in the exchange of records. On the other hand, FINMARC had advantages that we were not willing to give up, most important of which were the ISBD punctuation and field 248. The result of the evaluation was to move towards MARC21 but to keep some of the local features. The new format is a hybrid of MARC21 and is called MARC21-Fin.

The software selection process was arranged according to the European Union rules of procurement. During the final phase, we carried out an extremely thorough evaluation, with system demonstrations, hands-on testing, site visits and reference research, negotiations with the developers of the systems and financial evaluations of the vendors. The goal was to find the most functionally suitable and the most economically advantageous system for the local databases as well as the union and national databases. The essential guideline in the selection process was a fair and objective treatment of all parties involved. Since every step was documented, we would have been able to reconstruct the process, should it have proved necessary.

When the different parts of the selection process were drawn together, Voyager, produced by Endeavor Information Systems, Inc., best fulfilled the criteria. Voyager was found to be a complete, integrated system that was finished in the essential, traditional functions needed by the libraries, but which however is being further developed to meet the new needs and changing technologies. It fits both individual Linnea libraries and the Linnea network well. Local services can be streamlined and their scope extended. But centralized services will also benefit from Voyager via its consortium-driven functions. Increased efficiency is largely based on improved networking, since Voyager supports both Z39.50 and ISO ILL.

Special attention was paid to the union catalog functionality of the four final candidates. The new-generation software was seen to offer several enhancements to a union catalog compared to the old one. For the catalogers, it is easier and faster: the union catalog is updated automatically, as the

system copies new records from the local databases according to the configurations of these databases. For the users it is more informative, since there are real-time links from the union catalog to the local databases, displaying the status of each item. With the help of another Voyager function, Universal Borrowing, the user will also be able to place a request on the item.

The selection process was coordinated by the National Library, but all the libraries were heavily involved in the process from the beginning, when the selection criteria and the RFP were compiled, through the evaluation and testing of the systems until the end, when the decision was made. The directors of the libraries made the final decision by unanimously accepting the proposal made by the National Library. Voyager was selected as the new system for the Linnea network, and the contract with Endeavor Systems, Inc. was signed on February 4,<sup>4</sup> 2000, after the rectors of the universities had also approved the decision.

#### 4 The Network Architecture

The next question was how many servers an optimal solution for the Linnea2 network would require. In the Linnea1 network there were 17 HP3000 servers for the 25 databases. The number of servers was never really discussed during the implementation of Linnea1 because of the limitations of the computer technology of the time. Times were different now, and the consortium license signed with Endeavor enabled the libraries to have any number of databases or server machines. Accordingly, we had a free hand to pick the best network architecture for the Linnea libraries.

How far can one go in centralization? The answer depends on three factors: the available data transmission network, the capabilities of the software and the state of the computer technology.

The Finnish Academic and Research Network, FUNET, has been a key factor in the Linnea network since the beginning. Without the reliable infrastructure

---

<sup>4</sup> Annu Jauhiainen, "A New Library System for Finnish Research Libraries Chosen," *Helsinki University Library Bulletin* 2000: 12–19.

provided by FUNET it would not have been possible to use Linda as a cataloging tool in the manner we have since the early 90's. FUNET network allows libraries from all parts of Finland to efficiently access Linda and other union catalogs located in Helsinki. During the last two years the network has not been down even once. Given the extremely robust architecture of the network and reliable maintenance organization (Center for Scientific Computing, CSC), there are good reasons to believe that the FUNET network will remain at least as reliable and efficient in the future as it is now.

A shared server is not feasible for a library consortium if there can only be one database on the server. The Voyager software allows in principle an unlimited number of databases on a single server. However, practical experience from other Voyager consortia made it clear that there should not be more than about 5-7 databases on a single server, since a large number of databases may require much time for Oracle and Voyager updates: it may take several days to update many large databases, and during the process all the databases must be closed. Fortunately this problem has disappeared in subsequent Voyager releases; it is now possible to update databases on a shared server one at a time.

However, if all databases are dependent on the same database application or hardware and operating system process, severe problems would have an impact on every library simultaneously. Fortunately, new server technologies make it possible to have a single server and still avoid this problem: there are servers that can be internally split into several logical (and physical) parts.

Both Sun and IBM, which were the platforms Voyager supports and therefore the only possible candidates for the Linnea2 hardware solution, can deliver cluster-like computers. The high-end models of both the IBM and Sun product family can be separated into logical parts called domains (Sun) or nodes (IBM). Each part has its own operating system process and dedicated hardware from network card to processors. To the operators and users, the server looks like a cluster of computers.

There were, consequently, no technical constraints on choosing the network architecture freely. The National Library was eager to find out whether centralization would save money. The idea was not fully accepted by all at first, for a few computing centers were reluctant to give up

maintaining their own server. Therefore, at the request of the universities three scenarios were analysed:

- centralized model; all databases placed on a single machine;
- semi-centralized model; 3–5 servers;
- decentralized model; the current number of servers.

Cost analysis was based on both purchase price and the total cost of ownership, calculated for five years.

After a thorough analysis of the various options, a decision was made to choose the centralized architecture and buy Sun E10000 as the server system. The decision to go for Sun was based on technical merit and price. Both Endeavor and Oracle use Sun machines as their development platforms; this fact was also taken into account. Large computers such as the Sun E10000 have been optimized for heavy duty database usage and are also very reliable. Our practical experiences have shown that E10000 is indeed a very reliable server. Application-level problems in Oracle or Voyager are far more common than server problems, although still rare.

The Linnea2 server is able to handle 1400 active users, or more than 5000 concurrent users, about twice as many as before, on 17 HP3000 servers running VTLS. Both Endeavor, which did the calculations for the hardware configurations, and we felt that an ample safety margin was needed in order to avoid performance problems.

Immediately after the server was chosen, the decision was made to outsource the maintenance of the new server to the Center for Scientific Computing (CSC), a non-profit company owned by the Ministry of Education. It hosts Finnish supercomputers and maintains the FUNET network. CSC staff have excellent UNIX and networking skills, and are therefore very well qualified to maintain the E10000.

We have good evidence for the claim that an unprejudiced approach to server architecture has enabled us to combine significant savings with important technical improvements. Being a consortium helps a lot: libraries buying systems only for themselves will not be able to utilize new technology with similar efficiency. It is easy to understand from this point of view why library consortia are becoming more common in the US and some European countries. Finland has been one of the pioneering countries

in this area, and our experiences from such cooperation are very encouraging.

## Aspects of Centralization and Decentralization

Analysis of Sun and IBM hardware and discussions with technical experts led us to some generic conclusions:

- There is a general trend towards centralization, which started in the mid-90s, in commercial companies that are more aware of costs than public institutions. Universities have been slow in reversing their current tendency to decentralize, since the purchase price of small servers is approaching zero. However, the ever-growing number of computers means that operating costs are growing fast. Badly managed UNIX servers have already caused security-related and other problems, and things may get worse if decentralization continues;
- Hardware vendors are reacting to centralization (server consolidation) by developing systems that make it easy to consolidate applications from a large number of existing servers into a much smaller number of large computers via 'internal clustering.' The Sun servers such as 4800, 6800, 10000, 12000 and 15000 and IBM RS/60000 SP are good examples of this trend. In the future we will see even more systems of this kind from Sun, IBM and other vendors. Naturally these machines will be substantially faster than current ones; another prerequisite for centralization;
- Hardware vendors are capable of, and willing to, offer bargain prices for large systems. For workstations and small servers, proportionate discounts will always be much smaller than for large systems. If list prices are used for estimating purchase costs, centralized solutions may seem to be expensive. However, if negotiations are successful, a centralized server may well become the cheapest choice; and
- Never forget to estimate the total cost of ownership. Buying a number of small computers may look like a bargain, but taking all costs into account may change the picture. There are a number of things to remember: maintenance costs (paid to the hardware vendor), license and support costs (to the software providers), operating costs, plus miscellaneous

costs such as floor space occupied by the system and the electricity consumed by it.

## 5 The Linnea2 Consortium

During Linnea1, the cooperation between the libraries was never formalized. Collaboration was based on mutual understanding, with the National Library as the central agency, giving guidelines and working as an intermediary with the library system vendor, and with the Ministry of Education as the financier of the acquisition of the system and of the implementation. In the Linnea2 project there was no central funding from the Ministry, as had previously been the case; instead, the universities had to find the money from their general budgets. In addition to having a single contract with the software vendor, the members of the Linnea2 Consortium became owners of hardware that they had to administer jointly. It was considered necessary to have a formal contract and bylaws to ensure that decisions, especially concerning money, were handled in a way all members had agreed upon.

After the software, hardware and hardware maintenance contracts had been signed, it was time to legally establish the Linnea2 Consortium. The twenty universities, the Library of Parliament and the National Repository Library are the founding members of the consortium. New institutions can join as associate members that can buy services from the Consortium and from the National Library. According to the bylaws, most decisions, especially those dealing with money, have to be approved by the General Council, based on consensus. The Steering Group consists of seven members. The National Library is the executive body, preparing all the matters for the Steering Group and the General Council and representing the Consortium in dealings with third parties such as the software and hardware vendors (Endeavor Information Systems Inc and Sun Microsystems Finland), the hardware maintenance organization (the Center for Scientific Computing – CSC) and the outside world in general, for example the media. The Library is also responsible for organizing and coordinating cooperation and communication within the Linnea network.

The Linda database is owned by the National Library. The Consortium is not legally or organizationally involved with Linda. However, the

Consortium libraries are the main contributors to Linda, as well as the owners of the shared hardware and the software license. Therefore the National Library feels that it is important to discuss matters concerning Linda openly with the Consortium and have its acceptance in major issues.

## 6 Implementing Voyager

The implementation of Voyager in the Linnea network took place in the summer of 2001. The process started in April, and all local databases were using the new system by the beginning of the academic year. The implementation in the local library databases was smooth, considering how complex the situation was with so many databases and so many parallel loads. Including test loads, altogether about 35 million bibliographic records were converted from one character set to another, one cataloging format to another and one library system to another. In addition to the bibliographic data, acquisitions and circulation data were also migrated. This required very careful planning, taking into account human resources in the libraries, at the National Library, at the server maintenance organization and at Endeavor. Furthermore, everything was also dependent on the hardware resources. Fortunately we had a powerful server, which is divided into five logical parts, each of which could be used effectively for simultaneous loads.

The biggest challenges in the implementation were the size of the conversions (15 million bibliographic records and 26 databases), the tight schedules, the different conversions, multilingualism and different character sets (Cyrillic and Scandinavian characters), and communication among all parties. Thanks to sharing a physical union catalog, the data of the libraries were relatively homogeneous, which helped the conversion process.

The schedules were made tight on purpose. When Linnea1 was built, it took several years to implement VTLS in all databases. That was possible at the time, because most functions were manual and could continue that way as long as the implementation was finished and the new system was ready to be used. The situation was quite different now. The changeover to the new system had to be planned carefully in each database to make sure

that the functionality and the services in the libraries could continue seamlessly. The main reason for the tight schedule was, however, the union catalog. The libraries were dependant on Linda for copy cataloging and ILL localization. We could not afford to cut that tie for a very long time. Therefore the strategy was to migrate the library databases first and then the union catalog immediately after that. However, this plan did not quite work out as expected.

Implementing Voyager in the Linda database was not as easy as in the local databases. The reason for this was the fact that Voyager Universal Catalog was planned for consortia which had not had a union catalog before, but the catalog was created from the participating local databases at the same time as the data were migrated from the previous system to Voyager. The dynamic links between the Universal Catalog and the local databases were created during the load. However, in our case the union catalog already existed: we had Linda, a union catalog that was in very good shape. There had been a lot of duplicate records as the result of the initial loads in the early 1990s, in spite of our sophisticated duplicate detection algorithm. Those duplicates had been cleaned up little by little, and by the time we were ready to start the Voyager implementation, all duplicates had been taken care of. We did not want to lose all the work that had been done over the years and start from the beginning again. Endeavor was willing to do some development for us to enable Linda to be migrated and the dynamic links to be built differently from other UC sites. This development work was, however, more complex and more time-consuming than Endeavor had anticipated, which caused unfortunate delays in the implementation.

As of the fall of 2002, the implementation is still not complete. Endeavor has finished the initial loads, but we are still loading to Linda the material that has been cataloged into the local databases since the VTLS system was closed down in the summer of 2001. We are about to start ongoing life with the UC, which means constant real-time updating from the local databases.

In a large project like this, with a great number of libraries involved, communication is vital to a successful outcome. Communication between the libraries and the vendor had to be organized and coordinated. Communication and cooperation among libraries was equally essential. In

Linnea2, there was one new partner in the communication triangle. The arrangement of outsourcing the server was a new challenge to all partners. CSC had not worked with library databases before. The fact that the operation of all academic libraries is dependent on the server being up and running continually from early morning till late at night has required changes in their thinking and daily routines. For Endeavor, our solution is novel as well, in spite of their large number of customers worldwide. There are some centralized Voyager systems, but not to this extent. The maintenance organization being separate from the libraries or universities was also unknown to them. The change has, however, been most significant for the libraries. Until now, all except two of them have had their own server, maintained by their own people or by the computing center of their own university. By the time the common server was chosen, the libraries were more than willing to give up the maintenance of their own hardware. The long implementation period gave us a good opportunity to learn what living with a shared server really means. Each library always has to remember that in every respect they are not on their own, but must take their fellow libraries in the same E10000 domain into account. One configuration error, such as too long a timeout period, may cause problems in all libraries sharing the same domain, in spite of the safety margin in the server resources. We have unfortunately had some problems, but these occasions have taught us valuable lessons, and all parties should now be aware of how to avoid such incidents in the future. However, the fact that libraries now avoid the trouble and cost of maintaining their own UNIX servers is a significant improvement. A big help in the new situation was however, the strong<sup>5</sup> tradition of cooperation in the library system area for more than a decade.

---

<sup>5</sup> Annu Jauhiainen, "Voyager Implementation in the Linnea Network", *Helsinki University Library Bulletin* 2001: 10–13.

## 7 How the UC Linda Works

The Voyager Universal Catalog (UC) is a physical union catalog with real-time links to holdings and item information from the contributing libraries. Bibliographic records are the core of the database. Each bibliographic record has an attached holdings record, or several of them, indicating which local library database is holding the title. If the same bibliographic title belongs to several databases, the same number of holdings records is attached to the bibliographic record.

The records in the Universal Catalog are deduped. The deduplication process occurs when records are loaded into the UC, based on the duplicate detection profile, which is up to the library to establish. Voyager's duplicate detection algorithm does not fulfill the needs we have in Linda. We need to be able to separate almost identical records where only e.g. record types, languages, etc. differ, but that is presently not possible. The basic philosophy of duplicate control of this system needs to be changed in order to make that possible. Neither does the merge function in bibliographic duplicate detection work as a proper merge function should. This feature will be enhanced in the near future.

The holdings records are generated and attached to the bibliographic records when bibliographic records are loaded into the UC database. The 014a field of a holdings record contains the identification, which links it to the associated bibliographic record in the local database. The 852b field indicates to which local library database the record belongs. The UC holdings record only functions as a pointer or stub record in the dynamic connection to the local libraries' databases. As a search result, detailed holdings and item information is retrieved in real time from the holdings and item records stored in the local libraries' databases.

During Linneal, the catalogers were actually working in Linda. They cataloged everything directly to the union catalog and then copied the records to their own local database. Now the workflow is the opposite. Nothing is supposed to be done directly in the universal catalog. Records are cataloged (or in most cases copied from Linda or from some other bibliographic utility) into the local database. The system takes care of the rest. The cataloger need not know anything else; all that has been taken

care of by the system administrator, who has set up the necessary configurations.

There are several configurations that must be set on the system side before any records can be loaded to the UC. The settings include definition of each local library that the UC server connects to for detailed holdings and item information, duplicate detection profile, bulk import rules, cataloging policy definitions and security setups.

Dynamic retrieval and display of holdings and item information requires certain configurations on the local library side as well, in order for servers to connect to each other. Database definitions and connection information have to be set up in each contributing library database. In addition, there are some policy issues that need to be discussed, e.g. decisions have to be made whether to exclude certain records from the UC load. For example, such records might be acquisition records for titles that have not been received yet.

Once the configurations are set on both sides—the Universal Catalog and the contributing databases—every change in any of the local databases is updated in Linda. Records can be added, deleted or modified, and the change is reflected in Linda. The ongoing updates are bulk-loaded to the UC on the basis of the schedule set in the configurations. The bulk load schedule has to be defined separately for each database. The loads can be carried out every ten minutes or once a day, or even once a week, or at any interval in between.

## 8 Universal Borrowing

Voyager's Universal Borrowing (UB) function provides a structure for unmediated, reciprocal borrowing in a Universal Catalog setting. It allows the libraries to use their collections in integrated circulation and share the patron data. According to its basic philosophy, UB is patron-initiated and unmediated. Patrons of participating libraries can request and borrow material from any library within the Consortium. The material can also be returned to any library. All transactions are tracked in real time and patrons can follow the status of their requests, loans and possible fines and fees through the Web interface.

The use of Universal Borrowing requires a fair amount of technical work, in other words, a lot of configurations in each participating database. However, the technical part is easy, in spite of all the work. The technology allows almost anything, as long as you have taken care of the necessary settings. It is the politics that is the hard part. A lot of political decisions have to be made in order to get a sensible and usable functionality. That naturally takes time and requires agreements among the participating libraries.

The Linnea libraries have in principle decided to implement Universal Borrowing. At present a few libraries are starting to test it, in order to see how it fits our workflows and customs. The general trend within the Consortium is to encourage resource sharing and to help the users to get the books they need as fast and as cheaply as possible, even if that will most likely change the guidelines used within inter-library lending. One strict rule has been that users are not allowed to order from elsewhere a book that is held by their home library or any other library within the same city. This will inevitably change because the system does not yet offer a way to check the local holdings before the request is sent to another library.

Simultaneously with the testing period, we are supposed to agree on the political issues. First, there has to be an agreement on which libraries will participate in reciprocal borrowing. Is it going to be all libraries together, so that requests may be sent to any library in the Consortium? Or is the National Repository Library going to be a unilateral companion to each library, in its role as the repository for all of them? Each participating library will have to decide whether it wants to exclude certain collections from this function, preventing access by other libraries' patrons. Each library also has to decide whether it is going to allow requesting and borrowing to all of its patrons, or only to certain patron groups. Libraries together have to agree on the blocking of patrons (when and for what reasons) as well as on fines and fees. They have to decide whether they want to collect overdue fines or any other fees, and how the fines and fees are handled. Sending books from one library to another means costs, as Finnish universities no longer have mailing service free of charge. Since requesting is unmediated, the result will at least initially be a lot of books mailed from one place to another. A lot of books will be requested and never picked up for loan. Who will pay the mailing cost when a book is

sent back to the library where it belongs? The only solution seems to be to make students pay the mailing costs. It is also anticipated that books will at times be returned to 'wrong' libraries, even when it is not a universal borrowing loan to begin with. It is simply handy for a traveling student to return a book to the nearest library. Who will pay for the mailing of those books?

So there are a lot of open issues to be solved before this functionality is ready for use in the Linnea network. However, it is a marvelous way to encourage resource sharing in the tight economic situation.

## 9 Linda and the Polytechnics

The Finnish polytechnic libraries are at present in the process of implementing Voyager. There will be 28 Voyager databases after the implementation is over by the end of 2003. The polytechnic libraries have been using various systems and have until now not cooperated in the library system field. Nor have they had a union catalog of their own. Their serials holdings are included in Linda, but not their monographs. Now, as their implementation is moving forward, they are facing the union catalog question.

The polytechnics have three options at least in theory: to use a virtual union catalog, to have a physical union catalog of their own, or to join Linda.

The virtual union catalog is a suitable interim solution during the implementation phase when there are only a few Voyager libraries among the polytechnics. Once all 28 databases are up and running, the load on the server would be too high. The polytechnic libraries followed the example given by the Linnea libraries and purchased a shared server for all of their databases. The server is configured for the 28 databases only, and simultaneous search on all of them would be too much for it to handle. The number of the databases would also cause difficulties in duplicate detection when, at the maximum, records from 28 databases were displayed.

A separate union catalog for the polytechnic libraries only is a noteworthy option that has to be considered seriously. Creating such a union catalog would be relatively easy. However, the main problem with

this option is the cost: it would be necessary to purchase a new server, since the shared server the polytechnics libraries now have would not be able to cope with the union catalog database. The libraries should also buy a Voyager UC license and establish a maintenance organization for their union catalog.

From the point of view of costs, adding the polytechnics libraries to Linda is an attractive choice. There are also obvious functional benefits. It is estimated that the polytechnic libraries have a relatively small number of titles that are not yet in the Linda database. So the number of bibliographic records would not grow much if the data from the polytechnics were loaded to Linda, whereas the number of stub holdings records would be comparative higher. The use of the database would not be affected significantly either, since the polytechnics are already using Linda for searching as well as copy cataloging. The centralized server of the Linnea consortium has the resources to accommodate the growth in the number of records and also the increased use. Besides, it is possible to expand the server by adding CPU and memory, should that be necessary.

If the polytechnics' data were added to Linda, the number of libraries contributing to the database would be more than double what it is now. Furthermore, the new libraries do not have the same experience of collaboration as the present Linnea libraries, and they do not share the same practice in cataloging, nor the same level of standardization. That would mean an increased need for support. The Database Services within the National Library, the former Automation Unit of Finnish Research Libraries, is maintaining Linda and supporting the contributing libraries. The unit would have to be strengthened with new resources. However, that would be an easier and cheaper option compared to establishing a completely new support unit, even in the case of a separate union catalog for the Polytechnics.

These three options for the union catalog are under discussion at present. It is expected that decisions will be reached at the beginning of next year.

## 10 The Portal Project

The National Library has started a project for procuring software for the National Portal and Digital Library. There are two separate procurements, one for the Portal and the other for the Digital Object Management Software. According to our vision, the national network will in the future consist of three modules: Integrated Library System (Voyager), the Portal software (application to be chosen) and the Digital Objects Management System (application to be chosen). These three applications will have to communicate and work seamlessly together, as well as with other applications, via APIs and using open standards, to the extent that the patrons will see a single service.

According to a definition established at the workshop “Portals: Is There a Role for Libraries?” at ELAG, the European Library Automation Group, Semantic Web and Libraries, Rome, 17-19 April 2002:

A LIBRARY portal is an application which allows one-stop-shop access/searching and discovery via a unified single-point interface to organized heterogeneous resources and enabling services to a pre-defined community (users).

In the Finnish Academic Network, we see the portal as a gateway to the library databases, the union catalog Linda and other national databases, electronic resources, and collections, as well as remote databases which may be open to anyone, or commercial databases licensed by FinELib, the National Electronic Library. As of 2002, FinELib licenses cover about 120 databases and approximately 8200 scientific journals. With the help of the portal, Linda will be part of a huge virtual union catalog that connects all databases the user wants to include in the search.

The portal software must enable efficient searching of remote databases via Z39.50 or other means; it must be possible to exchange patron data between applications using the NCIP protocol and/or application dependent APIs, and all systems must support OpenURL for context-sensitive linking. OpenURL will have a direct impact on cataloging into Linda, for it is expected to solve the difficulties in maintaining the URLs of the electronic

---

<sup>6</sup> See <http://www.ifnet.it/elag2002/workshop.html>.

journals that are cataloged to Linda. But this is only one issue in the field of electronic material. Discussions on how to handle all electronic resources in Linda have only just started.

The procurement for the portal software is at the final stage. The decision will be made at the end of October. The plan is to implement it in a very short timeframe and be in production at the beginning of next year.

## 11 Conclusion

The Linnea libraries have been using the physical union catalog for nearly ten years. We have strong evidence of its advantages. We did not want to have a separate union catalog, the maintenance of which would require extra work. The aim has been, since the very beginning, to have a union catalog that is integrated into the local catalogs, in order to save resources in cataloging and to ensure homogeneity of data. The aim was already reached during Linnea1 and the first joint system. Linnea2, and the new generation system gave us a union catalog that is linked to the local databases in detailed holdings information and offers its users a lot of functional advantages. The next step will be a union catalog that will be a portal to the entire library network, and the basis for new services.

## **Chapter 6**

# **Beyond Technology: Power and Culture in the Establishment of National Union Catalogs**

Nadia Caidi

The purpose of national union catalogs (NUCs) is to facilitate access to the holdings of libraries in a nation, and to ensure that these resources are identified and easily located by a variety of users (scholars, students, general public, foreigners, etc.). National union catalogs are also useful in that they usually include the national bibliography of a country (i.e. grouping and recording of the publications of a country or about the country), although they go beyond the national bibliography to include holdings of other libraries and sometimes even records from large bibliographic utilities (OCLC, RLIN, etc.).

Union catalogs have developed in response to the need for library cooperation and resource sharing. By banding together and joining efforts to create a shared cataloging system, libraries in a given country (or across countries) create a foundation for resource sharing that reduces duplication of resources and cuts costs thanks to economies of scale.

Union catalogs' architecture can take various forms: physical, virtual, distributed, centralized, etc. (Husby 1999, Coyle 2000). The decision about which model of union catalog is more appropriate for a country's libraries is one that has to be made collectively by those engaged in the process. A range of players are usually involved in developing a union catalog; this includes the various libraries that contribute their records and the list of their holdings, but also other players such as system vendors, state agencies in charge of the various types of libraries, university administrations, funding sources, users and so on, all of whom may have their own agendas.

The NUC emerges as a result of the interaction between these different players; it becomes an artefact that is socially constructed by people who have a stake in its development.

As is often the case in any human activity that involves interpersonal relationships and negotiation, the process of developing a national union catalog is not devoid of tensions. The composition of—and the dynamics between—the various players in a country has a direct effect on the final outcome (i.e. the union catalog), and is key to understanding the choices made about the union catalog's design, architecture or functionalities. Much can be learnt from the negotiation process about the players involved and the power relationships between them, as well as the set of beliefs, values and practices that inform their decisions. The aim of this chapter is to raise awareness about the broader societal contexts that shape the establishment of union catalogs, with special emphasis on issues of power and culture.

## 1 The Social Shaping of NUCs

In an attempt to explore the social shaping of the NUCs and the negotiations around this artifact, a comparative study of the development of national union catalogs was undertaken in seven countries: Czech Republic, Slovakia, Hungary, Poland, Estonia, Latvia and South Africa (see Caidi 2002, forthcoming). The national union catalogs investigated were:

- NUKat, the Polish National Union Catalog;
- MOKKA, the Hungarian Shared Catalog;
- CASLIN, Union Catalog of the Czech Republic;
- the Slovak Union Catalogs of Periodicals and of Monographs;
- the Latvian Union Catalog;
- ESTER, the Estonian union catalog;
- SACat, the South African union database.

A mixture of face-to-face and structured telephone interviews (with follow-up by email) was undertaken with two rounds of data collected in 1999 and 2002. In-depth interviews were conducted with library directors, deputy

directors, heads of consortia, and project managers from the major academic and research libraries (including national libraries, academic libraries, state special libraries, central university libraries and other specialized research libraries). Although the library community encompasses a wide range of library types and sizes, the focus of this study was on the major university and research libraries because these libraries have been the most active in implementing library automation and information policies in their countries. Their involvement and collaboration was key to the success of the national union catalog projects. Interviews of those people who contributed their vision to and participated in the decision-making process allowed for a rich, complex and realistic picture of the social shaping of a NUC. A survey was also sent to union catalog coordinators and/or managers in each of the seven countries, in order to collect data on the architecture, functionalities and organizational aspects of the NUCs.

At the heart of the study is the idea that the development of these seven NUCs followed different trajectories based on the nature of the relationships between individual libraries in the country. The main question investigated was how much of the development of national union catalogs was influenced by differing visions and cultural practices, by the varying social contexts of the libraries, and by any personal tensions that may have contributed to the negotiation process over the union catalog (Caidi, 2002, forthcoming).

## 2 NUC Development in the Countries Studied

The seven countries studied all started their union catalogs in the mid- to late 1990s (in the case of South Africa, the union catalog was initiated in 1983 but was substantially revamped in 1997) and while the NUCs are at various stages of development, they are all operational, thanks to funding and support from The Andrew W. Mellon foundation, state agencies and other sources that made these nationwide initiatives possible (e.g. Soros' Open Society Institute, European Union funding, etc.).

These countries are obviously different, but they also have many elements in common. All have undertaken major socio-political transitions

(from the socialist regimes in Eastern Europe and the Baltic countries; and from apartheid in South Africa). Libraries, like all other institutions, were impacted by the changes and the resulting turmoil and uncertainties. These countries also received help and funding from foreign agencies, including western library-oriented philanthropic foundations, which provided them with much-needed funding and expertise in various areas, particularly as it relates to library development and automation (Borgman 2000; Lass and Quandt 2000; Quandt 2002).

Libraries in Central and Eastern Europe and the Baltic region—all governed under a socialist regime—have traditionally depended on different ministries, and were governed under a system that had aspects of both centralization and decentralization (Borgman 1996, 1997, 2000; Lass and Quandt 2000). The National Library (along with the network of public libraries) has traditionally been governed by the Ministries of Culture in the countries studied. Major university and research libraries were under the purview of the Ministries of Education, while other ministries (e.g. Agriculture, Science and Technology, etc.) were responsible for the various special libraries. State agencies and other institutions in charge of the various types of libraries, which are still very present in the governance of libraries, made it very difficult for libraries to undertake meaningful changes in their working styles.

In South Africa, the situation of libraries today reflects their apartheid legacy. Under apartheid, governance at all levels was based on racial lines, including the educational systems and library services. This situation led to vast inequalities between the relatively privileged white institutions and the much less privileged black, colored, or Indian institutions. After the transition, there were calls for major reforms of the higher education system, and in particular for a merger of these various institutions in an attempt to reduce duplication of resources and of the curriculum, cut costs, and allow for a more egalitarian educational system. In practice, however, the transformations have been slow and difficult to achieve. These changes have implications for the major university and research libraries. Cooperation had been taking place between libraries before the socio-political changes, although consortia tended (and still do) to form along local or provincial lines (e.g. the Cape province, Gauteng, etc.). The situation of South Africa is also different from the six other countries in

that SACat, the national union database, is run and managed by SABINET, a not-for-profit arm of Sabinet, Inc. (see chapters by Man and Erasmus and by Malan in this book).

Cooperation always existed in one form or another in the seven countries studied, but became the focus of library restructuring since the socio-political changes, partly because of the budget cuts to the cultural sector, and partly because funding sources favored projects that would benefit many libraries in the country rather than a few individual ones. The result has been an increase in inter-organizational linkages, attempts to adopt common standards and formats, the establishment of consortia and alliances, and the creation of shared cataloging systems.

Union catalogs are inscribed in this trend; they developed either as a result of a consortium of libraries that chose to use a common integrated library system (e.g. VTLS, Dynix, Aleph etc.), or that banded together because of close geographical proximity (e.g. GAELIC, FRELICO, CALICO etc. in South Africa). It was only a matter of time before libraries in the countries studied deemed it necessary to create a national union catalog and sought funding to develop it (more accounts of the origins and initiation of the different NUC projects can be found elsewhere in this book. See also Caidi 2003, forthcoming; Quandt 2002).

**Table 1. NUC Projects' Initiation**

Union Catalog	Project Dates
CASLIN UC (Czech Republic)	Start of Project: 1999–2000; Operational: 2002
Union Catalog of Slovakia	Start of Project: 1999–2000; Operational: 2002
MOKKA (Hungary)	Start of Project: 1997; Operational: 2002
NUKat (Poland)	Start of Project: 1997; Operational: June 2002
Union Catalog of Latvia	Start of Project: 1997 ; Operational: March 2000
ESTER (Estonia)	Start of Project: 1995 ; Operational: January 1999
SACAT (South Africa)	Start of Project: 1983–84; Revamped: 1997

With the help of funding agencies eager to see libraries work together to develop shared cataloging systems and make their holdings records available online for all to access, national union catalog projects took off (see Table 1).

### 3 Beyond Technology: Power and Culture

Technological artifacts can be viewed as socio-technical systems in that they involve more than just the solving of technical or design problems, but also include the overall dynamics that contribute (or do not contribute) to their development. Agreeing on the terms of the collaborative endeavor is a complex process that includes elements of power and culture. These two elements are essential in understanding the dynamics at work in negotiating the different stages of the development of NUCs.

#### Power Issues

The Merriam-Webster dictionary defines power as “possession of control, authority, or influence over others” or the “ability to act or produce an effect.” Power usually stems from interacting with others. Through interpersonal relationships, various attempts—conscious or not—are made by an individual or a group to impose their views or will over others. Methods of doing so include persuasive arguments (moral or financial), emotions, reason, etc. At the heart of power is control or influence over the outcome or the process. A clear delineation of the goals and objectives and how to reach them (sharing of responsibilities and delegation of power) is, therefore, critical in any cooperative endeavor. Communication and trust are key to negotiating the power balance.

When asked to reflect on the lessons learnt from establishing their NUC, three factors were deemed essential by respondents in most countries and can be summarized as follows: the technological aspects; the organizational aspects and the vision or ‘philosophy’ about what the NUC should be.<sup>1</sup>

---

<sup>1</sup> For a more extensive discussion of these findings, see Caidi 2002 and Caidi, forthcoming.

Technological aspects concern the choices made about the overall architecture of the union catalog, including the library system, the standards and protocols adopted, the cataloging procedures, the content and functionalities of the union catalog. The philosophical aspects—a term used during the interviews by a few respondents—relate to the visions (shared or not by the various libraries) about what the union catalog should accomplish and how it fits within the overall information infrastructure of the country. Finally, the organizational aspects address such issues as who will build, operate, maintain and finance the union catalog (Coyle 2000; Husby 1999; Lynch 1997).

At various stages of their development, the national union catalogs studied have had to deal with issues of power. At the planning stage, issues of power translated into numerous questions: Which formats should be used? Who makes the decisions? Whose interests are taken into account? What is the influence of the type of technology? What are the characteristics of the players and their network? and so on. A national union catalog is a product of the decisions made by the collective group: decisions about what to centralize and what to decentralize have impacts on the sharing of labor and the sharing of responsibilities, and contribute to defining or shifting the power balance. In all the countries studied, there was an interest in achieving a balance between centralization and decentralization. However, in practise the existence of interest groups, ‘fan clubs’ of various integrated systems, and alliances of various natures made for a far more complex picture and led to various tensions. Because union catalogs emerge as a result of the ‘work’ of many types of actors, the variety of coalitions influences the development of the artifact and leads to different visions or ‘philosophies’ of what the artifact should be and how it should be designed and used. When a socio-technical system involves the cooperation between various groups, each with its own understanding and conception, there is room for miscommunication or conflicts.

The issues of power also arise at the development and maintenance stages of the NUC. The organizational communication and management literature refers to this stage as the ‘commitments’ stage (Ring and Van de Ven 1994), or what Kanter (1994) in his marriage analogy divided into the “getting engaged,” “setting up housekeeping,” and “learning to collaborate” stages. Indeed, much like a marriage, without maintenance a technological

system becomes a ruin. The questions that are raised at this stage therefore include: Whose interests prevail? Those of the designers of the system, the end-users, the funding sources, the IT people, the library staff, etc.? What are the incentive mechanisms? How do conflicts get solved? Who plays what role (e.g. change champions, mediators, agitators, etc.)? How is trust achieved? How open and transparent is the process? A classic problem in cooperative work is the trade-off between individual and collective goals. Hofstede (1980) uses this dichotomy in his study of cultural variability; according to him, certain cultures place the emphasis on collective socio-economic interests over those of the individual. On the contrary, individualistic trends in other cultures override any collective attempts (i.e. everyone is expected to look after themselves and their own needs). Such a dichotomy was observed in the study and was present in various degrees across the seven countries. The implications for the establishment of a NUC are manifested in the willingness or unwillingness to share information among members of the consortium, whether access to the data was restricted to members of an institution versus users from other institutions, and the extent to which tensions resulted from personal ambitions and interpersonal conflicts (see Caidi 2002 and Caidi, forthcoming, for examples from the seven case studies).

## Cultural Issues

Culture has been defined in different ways. It usually refers to the system of beliefs, attitudes and values shared by members of a group, whether it be at nation or country level (the sense of belonging to a particular nation or ethnic group); at domain level (bonding based on expertise, areas of interest or specialization); or at the level of the organization (loyalty and mores shared by members ('insiders') of an organization). In some regard, it may be more relevant to talk about 'cultures' (or 'identities') rather than the generic term 'culture,' because one can belong or identify with various communities and at various levels. Cultural aspects are used here to refer both to the choices and values that are embedded in the design and use of information, its agencies, and its technologies; as well as to how these might translate across cultural contexts.

The findings show signs of this trade-off between globalization and localization. The countries studied are at various levels of their transition from their earlier regimes to a democratic society with a liberalized market. The data make references to the attempts to balance the need for global integration into the world economy with the preservation of the local language and cultural identity. One clear example is around the discussion over the adoption of standards. Most countries have adopted the major library standards (AACR2, MARC21, ISBD, UDC, LCSH, etc.), although some have maintained the local variant (e.g. Hungarian or South African versions of MARC, Polish standards for bibliographic description and classification, Estonian Universal Thesaurus, etc.). While the adoption of standards ultimately allows libraries to exchange data with libraries throughout the world and join the international library scene, the adoption of these standards requires adjustments or radical changes to existing practices. Some people may resist those changes, and frame the debate along cultural imperialism lines (e.g. wishing to maintain one's cultural uniqueness).

Most respondents in the countries studied viewed the union catalog as a means to open up a window on the world and disseminate their country's rich literary heritage. Others, however, were cautious, pointing out that it was essential that their language and cultural identity were preserved and adequately protected. Hungary, for instance, is an island amidst the Slavic countries, with a distinct culture and language that it ferociously seeks to preserve. Similarly, Slovakia, which became an independent nation in 1993 for the first time in its history, is busy creating its national identity and preserving its language and cultural heritage. It is only recently that South Africa has been considering converting to MARC21 (formerly USMARC). Until then, the South African version (SAMARC) was the most prevalent among libraries.

The development of a NUC therefore presumes a few key assumptions:

1. Technology and its use are part of a culture;
2. Biases are often embedded in tools themselves (system architecture, modules and functionalities; templates, icons, organization, computing metaphors, etc.); and

3. Linguistic issues go beyond the translation of the commands in the local language to include organization of the knowledge, cultural constructs, representation, metaphors, etc.

In summary, the findings point out that no country fared better than any other. Rather, issues were strikingly similar across countries and differences in the outcome usually had more to do with the group size and group boundaries, the dynamics between members of the group, the incentive system, and the support received (or not) at different levels. The study of the NUC across these seven countries also points out the increased awareness of the dynamics and mechanics of cooperation, the pivotal role of communication; as well as the importance of good leadership and accumulation of local knowledge.

#### 4 What is Next for NUCs?

Beyond the design and development stages is a critical test for the technological artifact: that of its usability and usefulness for all relevant users. Questions that arise include: How easy is it to figure out and learn? How efficient is it (e.g. requiring as few steps as possible to retrieve desired information)? How easily can steps be remembered? How can one make sure that the NUC is used? How can one assess its usability and usefulness?

These questions are essential to determine who will use the NUC, and how. The concept of usability is predicated on establishing criteria for effective, efficient and satisfying use, and it is certain that cultural variability plays an important role in determining such criteria. As yet, usability practitioners have rarely articulated this issue. Previous research on the use of online public access catalogs (Duncker 2002) and Internet search tools (Iivonen and White 2001) have shown differences in how users from different cultural groups search for information.

Culture, as Hofstede (1980) puts it, is a collective mental programming. Like any socio-technical system, a union catalog embodies the values, beliefs and practices of its producers, along with their broader social and cultural contexts. A user with different sets of beliefs and assumptions about the organization of the content, the categories assigned or the user interface design may find it hard to interact with the system. Lessons

learned from cross-cultural usability and international user interface design are thus important for the design of usable NUCs (Caidi and Komlodi 2003).

What is increasingly needed is more research on information-seeking behavior of users in transitional societies (or in general of user studies outside North America and Western Europe).<sup>2</sup> Some of the countries studied have had a long history of central planning and an information culture that promoted a particular form of interaction with knowledge, as well as learning styles that emphasized memorization over critical thinking and independent research. The implications for libraries were that the priority was on building collections rather than providing services to users. As a result, very little attention was paid to end-users' needs and their seeking behavior (e.g. explicit behaviors (search strategies used, evaluation of particular resources, problem-solving, etc.) as well as implicit cognitive models, categories and metaphors). The aim should be to enable the design of systems that cater to individual differences and various cognitive models. From a cross-cultural usability perspective, there is also a line of research that could look at the operationalization of culture for the purpose of enhancing usability as a means to assess whether culture is a significant variable in usability design (Caidi and Komlodi 2003). Research on the internationalization of industrial products, software or websites exists, along with an increasing interest in research on interface design for multi-cultural environments. However, relatively little research exists on these issues in the literature on library information systems.

## 5 Concluding Remarks

While technologies may be global in nature, their use, content provision and design have remained local. The study briefly outlined in this chapter,

---

<sup>2</sup> Exceptions include a study of information-seeking behavior of Mongolia's urban residents (Johnson, 2003) and a conference on "Information Behavior in Digital Libraries" held in Bratislava, Slovakia, on May 21–23, 2003.

as well as the discussion above, point to the question of appropriation or acculturation of the union catalog and its subsequent use by various groups.

In order to investigate the 'acculturation' process in the context of the library scene, one needs to examine the ways in which a technological artifact is appropriated in various cultural milieux. In other words, how does a technological artifact become 'localized' and used by various groups who may be the intended audience but who were not the designers and/or developers of the technological artifact? The findings are particularly interesting in the context of the situation of libraries in transitional societies, where transfer of technology was made possible through various organizations (e.g. philanthropic foundations, non-governmental associations, etc.). There are both exogenous and endogenous forces that contribute to adoption and use of information technologies, and the extent to which foreign agencies and philanthropic foundations shape the development of information infrastructure in a given country is a critical issue (Caidi 2003).

There is no question that philanthropic foundations, state agencies and other funding agencies have vastly contributed to these nations' information infrastructure by providing them with the funding and technology needed to improve their libraries and automating their internal and external processes. However, beyond the technology transfer, various forms of knowledge transfer also took place which will allow the libraries in the country to build or rebuild their social capital, to provide training in the form of seminars and workshops on cooperation and resource sharing, and to allow local knowledge to accumulate. It is time for information scientists to address these important questions and raise awareness about the need for research in the area of usability of union catalogs (and digital libraries in a broader sense) and user studies. The goal is to identify new tools, techniques and methodologies for cross-cultural study of user behavior in digital libraries and international user interface design, and to provide a forum for generating new research directions and cross-disciplinary collaboration.

Libraries as social and cultural institutions have much to contribute to the development of their country's information infrastructure. After having integrated library systems and developed their online public access catalogs, libraries are coming together to solve common issues, to serve the needs of their users and contribute to the development of NUCs and digital

libraries at national level. The question that remains unaddressed for the NUC is how to make it a part of the broader national information policy? Indeed, how to ensure that libraries and librarians play a key role in the policy arena in their country? The free flow of information in a society is as critical as the political and economic reforms or technological advances. It is essential that the library communities organize themselves and use the lessons learnt from developing a national union catalog to form new collaborative alliances that will enable them to remain actively involved in the development of a national and increasingly global information infrastructure.

## References

- Borgman, C. L. *From Gutenberg to the Global Information Infrastructure: Access to Information in the Networked World*. Cambridge, MA: The MIT Press, 2000.
- Borgman, C. L. "From Acting Locally to Thinking Globally: A Brief History of Library Automation." *Library Quarterly* 67/3 (1997): 215–249.
- Borgman, C. L. "Automation Is the Answer, But What Is the Question? Progress and Prospects for Central and Eastern European Libraries." *Journal of Documentation* 52/3 (1996): 252–295.
- Caidi, N. "Cooperation in Context: Library Developments in Central and Eastern Europe." *Libri* 53/2 (2003): 103–117.
- Caidi, N. "The Social Shaping of National Union Catalogs: A Comparative Analysis." Paper given at the Mellon Foundation Conference on National Union Catalogs, Tallinn (Estonia), October 17–19, 2002. Available electronically at <http://www.nlib.ee/inglise/docs/mellon/melprog.html>.
- Caidi, N. *The information infrastructure as a discursive space: A case study of the library community in Central and Eastern Europe*. PhD dissertation, Department of Information Studies, UCLA.: 2001. (Adviser: C.L. Borgman).

- Caidi, N. (forthcoming). "The National Union Catalog as a Boundary Object: A Comparative Analysis." To appear in *Library Quarterly*.
- Caidi, N. and A. Komlodi. "Cross-Cultural Considerations in Digital Library Research." *D-Lib Magazine* 9, no. 7/8 (July/August): 2003. Available electronically at <http://www.dlib.org/dlib/july03/07contents.html#IN-BRIEF>.
- Coyle, K. "The Virtual Union Catalog." *D-Lib Magazine* 6, no. 3 (March): 2003.
- Duncker, E. "Cross-Cultural Usability of the Library Metaphor." *Proceedings of the JCDL '02 Conference, 2002*: 223–230.
- Hofstede, G. *Culture's Consequences: International Differences in Work-related Values*. Newbury Park, CA: Sage: 1980.
- Husby, O. "Real and virtual union catalogs." *Proceedings of the CASLIN'99 Seminar on Union Catalogs: Their Organization and Services*, May 30-June 3, 1999. Available electronically at <http://www.caslin.cz:7777/caslin99/a2.htm>.
- Iivonen, M., White, M. D. "The Choice of Initial Web Search Strategies: A Comparison between Finnish and American Searchers." *Journal of Documentation*, 57/4(2001): 465–491.
- Johnson, C. A. *Information Networks: Investigating the Information Behaviour of Mongolia's Urban Residents*. PhD dissertation, Faculty of Information Studies, University of Toronto: 2003.
- Kanter, R. M. "Collaborative Advantage: The Art of Alliances." *Harvard Business Review*, 27/7–8(1994): 96–108.
- Lass, A., and R. E. Quandt (eds.). *Library Automation in Transitional Societies: Lessons from Central and Eastern Europe*. New York: Oxford University Press, 2000.
- Lynch, C. "Building the Infrastructure of Resource Sharing: Union Catalogs, Distributed Search, and Cross-Database Linkage." *Library Trends* 45/3(1997): 448–61.

- Quandt, R. E. *The Changing Landscape in Eastern Europe: A Personal Perspective on Philanthropy and Technology Transfer*. New York: Oxford University Press, 2002.
- Ring, P. S. and Van de Ven, A. H. "Developmental Processes of Cooperative Interorganizational Relationships." *Academy of Management Review*, 19 (1994): 90–118.



## Part 2

### Czech and Slovak Union Catalogs



# Chapter 7

## The CASLIN Union Catalog

Gabriela Krčmařová and Ilona Trtíková

Virtual union catalogs revolve around cooperating technologies, real union catalogs revolve around cooperating people.

### 1 Introduction

The CASLIN Union Catalog (Union Catalog or UC for short) is a centralized national union catalog. It is a single database that collects documents stored in Czech and Moravian libraries, which use a variety of library systems. Since 2000, it has operated in a tailor-made system called CUBUS, designed to fully meet the requirements of maintaining and operating exactly this type of union catalog. The launch of the Czech National Union Catalog was one of the tangible results of the CASLIN Project (Czech and Slovak Library Information Network) [17]. Between 1993 and 1995, the CASLIN Project gave life to all activities important for a national union catalog. Besides a clearly defined and detailed concept (modeling a union catalog in operation, gathering and maintaining data, user categorization, etc.), fundamental standards were established, and the Union Catalog administrator was identified.

In July 2002, the Union Catalog contained 1,578,868 records of printed monographs and special types of documents from 110 libraries. There were 60 active members regularly supplying the Catalog with records. In addition, the Catalog contains 84,683 records of serials from 550 participants. The directory of libraries and information institutions contains 2,947 records.

## 2 The Fundamental Strategy and Standards for the Union Catalog

The undeniable advantage of having to bridge a 20-year gap to the Western world lay in the opportunity to determine unified standards for record provision and exchange before launching the National Union Catalog (1995). However, this was the only advantage. The National Union Catalog was established in a similarly heterogeneous environment to that of its Western counterparts 30 years ago [23].

The following library systems existed in the Czech Republic:

- Academic libraries have been using TINLIB as their integrated library system since the early 90s;
- Public libraries have mostly started to implement a Czech library system called LANius, or later Clavius;
- The Czech National Library and other large libraries implemented the ALEPH integrated library system.

The heterogeneous nature of the library environment is further reinforced by other library systems, both Czech and foreign: KpSys, Rapid Library, Olib, ISIS and WINISIS, Daimon, etc.

The strategy of the Czech Union Catalog is the same as the fundamental conceptions of major union catalogs throughout the world. As a matter of principle, the Union Catalog is open to all libraries and information institutions able and willing to respect the established standards of record provision:

1. The primary exchange format is UNIMARC Exchange Format, with CDS/ISIS as secondary exchange format;
2. ISBD(G) is the basic standard for name processing;
3. The guiding rules are AACR2, 1998 edition;
4. UDC notation is binding for subject cataloging;
5. The Union Catalog Record determines the binding record format for particular document categories, as established by the administrator of CASLIN Union Catalog ČR. In its Standardization Series, the Czech National Library published the following union catalog instructions:
  - Union Catalog Record: UNIMARC. Printed Monographs (1996, Standardization Series #4);

- Union Catalog Record: Exchange Format. Printed Monographs (1997, Standardization Series #9);
- Union Catalog Record: UNIMARC. Special Types of Documents (1999, Standardization Series #16);
- Union Catalog Record: UNIMARC. Printed Serials (1999, Standardization Series #17).

From the very beginning, the Union Catalog has focused on general technical standards: TCP/IP, HTTP, FTP, etc.; the implementation of Z39.50 communication protocol has been planned for 2002.

### 3 The Functionalities of the Union Catalog

Currently the Union Catalog serves the following objectives [26]:

#### Information Function

The Union Catalog is a source for searching for and finding a particular document, or gathering information about documents concerning certain topics.

#### Document Location

The Union Catalog allows the user to locate the library that holds the document in question, and possibly also to obtain detailed data about the document's shelfmark, usually facilitating the borrowing of the document.

#### Document Retrieval

The Union Catalog makes it possible to act on a request to borrow a document or request its copy (Inter-library Loan Service—ILS). This service does not necessarily have to be a part of the Union Catalog. However, this type of service is often offered together with the ability to decide to which library this request should be forwarded, while keeping in mind the possibilities of any given library.

## Shared Cataloging

The Union Catalog is a tool for shared cataloging, offers access to records and their copying, and is a tool for the formation and optimization of name and subject authorities.

### 4 Union Catalog Services

The fundamental principle in creating a union catalog is the controlled harvesting of data of the broadest possible scope, with the aim of creating a concentrated information base and a qualitatively and quantitatively rich source of secondary documents (records). This principle, if followed, allows for the introduction and development of additional services for the users of libraries and information institutions, as well as for librarians themselves [26].

The Union Catalog offers the following services:

- Searching, i.e. locating documents in Czech libraries;
- Provision of reference and inter-library services, i.e. sending a loan request to ILS, where the identifying data for a library are generated from the directory and the document data from the record in the Union Catalog;
- The Clipboard, used for storing located records for later use (printing, data export);
- Receiving document records for retroconversion of local library catalogs, for current cataloging, or for the national bibliography;
- Shared cataloging, in order to process current production in two ways: copy cataloging, and online shared cataloging by means of a preset input form; and
- Use of the input form to edit data in existing records and in the location data of the member library that owns a given document but does not send its documents to the Union Catalog.

The Union Catalog users are subdivided into three categories, based on the type of services they use. There is a fundamental prerequisite for any union catalog to achieve its goals—it must be filled with data.

## 5 Union Catalog—Data Administration Maintenance

In the Czech National Library, the Union Catalog Department is in charge of the Union Catalog's administration. Libraries mostly contribute to the Union Catalog offline, i.e. from time to time they upload batches of newly processed records or records formed in the retroconversion process.

In January 2000, the Union Catalog began to operate in the CUBUS system, and all processes involved in union catalog database administration, including input data analysis, were automated, making maximum use of existing software tools [19]. Record processing is automatic, and its steps are

- receipt and identification of a data file (including conversion);
- formal logical data control: a UNIMARC test and a duplicates test;
- data import;
- statistics for the participants and the administrator, and
- problems left for the administrator's decision.

A participant library places its data in its allocated space on the FTP server (if the data are delivered on a floppy disk, they are transferred to the FTP server by the administrator). The program periodically checks whether there are new data on the FTP server. If so, the program downloads the data and, using the name convention (see below), identifies their owner, their format, and the character set used. For import, it is important to name the file in compliance with the naming convention.

### Name Convention (Name Format)

The data filename may have up to 8+3 characters (i.e. 11 in total). The first 6 characters identify the institution, characters 7 and 8 stand for the code page used, and characters 9, 10 and 11 identify the data format (e.g. aba001kg.vfi).

**Table 1. Character Set (Characters 7 and 8)**

um	ISO 646 or ISO 5426
gi	all accent marks are recorded by means of the GIZMO notation
lg	PC Latin 2 (Microsoft Code Page 852) + GIZMO
kg	Code Page Kamenicky + GIZMO
uc	UNICODE UTF 8
sg	ISO 8859-2 + GIZMO
an	ANSEL

**Table 2. Data Format (Characters 9 through 11)**

dat	file exported from ALEPH
rum	textrow UNIMARC
uis	UNIMARC ISO 2709
vfo	ISO 2709 exchange format
vfi	exchange format, file exported from the CDS/ISIS system
dt	file exported from ALEPH 500

An example is provided by aba006lg.uis, where the institution identifier is ava006, the Code Page is PC Latin2 + GIZMO and the data format is UNIMARC ISO 2709. All records, both the new and the edited ones, are tested before they are imported into the Union Catalog.

The automatic test is set up

- to test the file for UNIMARC compliance: an application automatically tests individual records for their compliance with UNIMARC field format;
- to weigh the record quality: there are six quality weight grades (4, 9, 10, 11, 12, 20), and the higher the grade, the better the record;
- to test the record for duplicates: the records are automatically tested for duplicates, and, if necessary, the duplicates test is accomplished in two

stages, the duplicate record with higher weight replacing the one with lower weight.<sup>1</sup>

Having tested a given record for duplicates in the current Union Catalog, the CUBUS system imports it into the Union Catalog database and then compares it with all the records within the currently processed batch, i.e. all records are also tested for internal duplication within their own batch. The resulting report on non-complying records is sent (by e-mail) for correction with appropriate commentary to the library that provided them. In case of a 5,000-record batch, the whole processing takes about 40 minutes and the administrator can set the start date and time. A member can also read the import result from the statistics available after entering the password at the CASLIN website.

## 6 The Array of Libraries Cooperating with the Union Catalog

As a matter of principle, the Union Catalog is open to all libraries and information institutions in the Czech Republic, regardless of the library system they are using. Libraries cooperating with the Union Catalog include

- central, universal, and specialized libraries;
- university and college libraries, and libraries of the Czech Academy of Sciences;
- public libraries in statutory and district towns;
- other libraries whose collections comply with the qualitative standards of inter-library services.<sup>2</sup>

---

<sup>1</sup> The UNIMARC testing algorithm, weight calculation algorithm and the algorithm for primary and secondary record comparison are described in detail in the document “CASLIN—Union Catalog ČR”. The data necessary for programming the ORACLE-based applications are available at the CASLIN website (URL <http://www.caslin.cz>)[13].

<sup>2</sup> The list of libraries contributing their records to the Union Catalog, including the number of records supplied, is available at the CASLIN website (<http://www.caslin.cz>).

The growth of the database shown through the number of records since 1995 is shown in Figure 1:

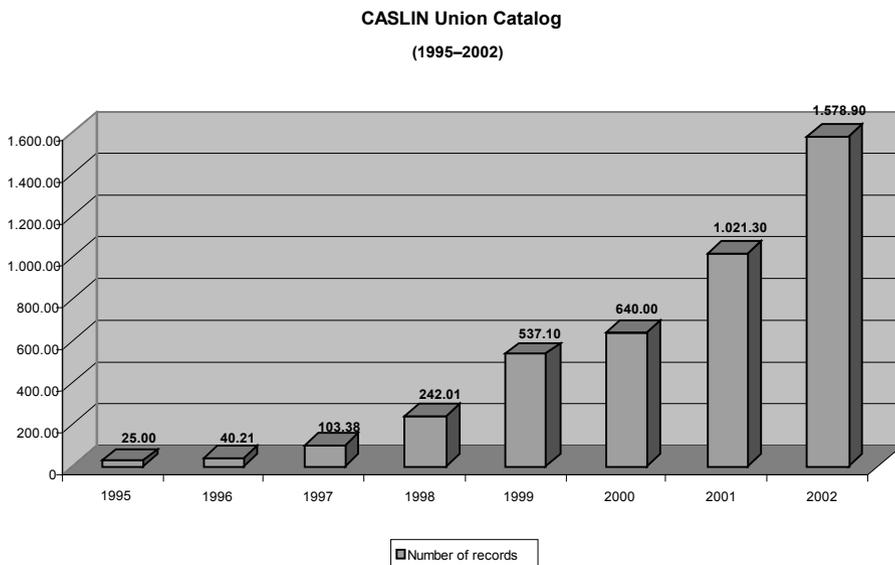


Figure 1. CASLIN Union Catalog, 1995–2002

## 7 Union Catalog Hardware

There are no special hardware and software requirements for the Union Catalog users. All services provided by the Union Catalog work well with the Netscape browser version 4.04 and higher and with the MS Explorer version 4.01 and higher. The Union Catalog is operated on an Alpha Server 1200 with 1GB RAM. Disk capacity is around 100GB. Digital Unix is the operating system used. The database server is provided by Oracle.

## 8 Union Catalog Software

Since 1995, the Union Catalog has been operated in three different systems: CDS/ISIS, ALEPH, and CUBUS.

### CDS/ISIS (1995–1996)

In the years 1995–1996, the Union Catalog was operated under the CDS/ISIS system as a union database with duplicates testing, containing records on foreign documents only (Union Catalog CEZL—*Centrální evidence zahraniční literatury*). The records were regularly converted to the ALEPH system and made available on the Internet, while only the search function was possible on the Union Catalog. In late 1996, the Union Catalog contained more than 40,000 monograph records.

### The ALEPH System (1997–1999)

In the years 1997–1999, the Union Catalog for monographs was operated in the ALEPH system and included domestic as well as foreign documents. Duplicates testing was available, although it was cumbersome and performed by external programs. The Union Catalog was used for searching (location of a document), and offline record sharing was possible only between the members that used the same version of ALEPH.

ALEPH made it possible to solve two key problems:

1. Online shared cataloging;
2. Duplicates testing, including the preservation of better-quality records and formal logical tests.

#### Problem #1: Online Shared Cataloging

Even earliest strategic analyses for the future union catalog maintained that “the target principle of the CASLIN Union Catalog is online shared cataloging.” However, ALEPH does not allow any manipulation of database records from outside itself, which, as a practical matter, prevents any online shared cataloging for members using different library systems, since they cannot carry out primary cataloging in the Union Catalog’s own

database [21]. This fact was officially communicated to the Union Catalog members at the Union Catalog Task Team meeting on June 3, 1997 [28].

### Problem #2: Duplicates Testing

ALEPH is a high-quality library system which, however, does not allow for the import of records with duplicates testing that matches the needs of a real Union Catalog operating in a heterogeneous environment. When it became obvious that Ex Libris was not able to satisfy the special requirements of the Czech National Library and modify the programs supporting Union Catalog administration (the ULM module), a Czech company developed a duplicates testing program. The program is of very good quality, but since ALEPH does not permit record manipulation from outside ALEPH, the necessary program modifications were somewhat cumbersome and had to be carried out outside the database. This solution meant that the duplicates testing procedure had to be activated outside the ALEPH database, and hence it was only possible to process data offline [3]. Another highly restrictive factor was that the duplicates testing and logical inspection procedure required the administrator to start seven support programs manually. The procedure for processing one batch of data provided by one library (regardless of whether the batch contained 100 or 10,000 records) took two days. And the time requirements were increasing in direct proportion to the size of the database. As a result, it was not possible to import the records supplied by the ever-increasing number of Union Catalog participants in real time [19].

These issues triggered a discussion on developing a system of administering and operating the Union Catalog with our own resources. In September 1997, the Union Catalog administrator presented a document entitled “A Potential Path of Future CASLIN Union Catalog Development with Regard to the Up-grade to ORACLE 7” [12], and in October 1997 we applied for a grant from the Mellon Foundation, which would allow us to purchase the ORACLE database system and develop our own union catalog system.

ORACLE became the tool for developing a system for the administration and operation of union catalogs, and the new system was called CUBUS. For system development, we chose a smaller software company with which we had cooperated since 1994 and which also created

the duplicates testing program mentioned above (fully implemented and improved in the new system). The company was paid by funds awarded by the Ministry of Culture of the Czech Republic to the project “CASLIN Union Catalog Development”. In November 1997, a representative of the software company presented the basic philosophy of the new system to members of the Union Catalog Task Team and to members of the Union Catalog Research and Development Team. Also invited were the directors of all major libraries and representatives from the Slovak Republic. The crucial task was to draw up the requirements for the development of a new system.

### The CUBUS System (since 2000)

During the first quarter of 1998, ORACLE was installed on our ALPHA server, and in March 1998, the “Requirements for Application Development under the ORACLE System” [13] was published on the CASLIN website. The members of the Union Catalog Task Team and of the Union Catalog Research and Development Team were invited to submit their comments to the Union Catalog administrator. After processing the comments in mid-1998, the development of CUBUS was launched. The resulting product contained the following improvements [17]:

1. Work procedures were optimized by eliminating human intervention wherever possible and effective;
2. Data control was improved by supplementing “human-based” control with automatic formal logical control;
3. The comparison keys for duplicates inspection were expanded and became more subtle, resulting in a reduction of unwanted duplicates in the Union Catalog [9];
4. Statistical monitoring was introduced to cover the movement of data, the administrator’s performance and the work of Union Catalog members;
5. The user interface appearance and functionality followed the access rights setup and configuration changes,
6. Shared cataloging and editing of old records already in the database was improved;

7. Data security was improved; and
8. The possibility was added to modify existing applications and to develop new ones for better performance in the future.

The beta version of CUBUS was provided to the Union Catalog Research and Development Team in September 1999 and to the Union Catalog Task Team in November 1999. Most of their comments were implemented and, early in 2000, the Union Catalog under the new system became accessible to the general public.

Since 2000, the Union Catalog has operated under CUBUS, which fully meets the requirements of the Union Catalog administrator as well as the needs of its users. CUBUS is owned by the Czech National Library. It is equally open to all Union Catalog members regardless of the system used by the member in question, and thus solves the problem arising from the heterogeneous nature of the library environment in the Czech Republic.

CUBUS offers solutions in critical areas:

1. Real-time data import in supported formats and code pages without compromising any further operations within the Union Catalog database;
2. Formal logical data inspection;
3. Duplicates testing using both primary and secondary keys with further amendments for series;
4. A search interface reflecting UC member requirements and capable of parallel processing of any number of simultaneous search requests from different users;
5. A direct link to the Library Directory is provided;
6. It distinguishes between users and active UC participants;
7. It provides batch-based data export in the supported formats and character sets;
8. It allows the user to place ILS loan requests;
9. It uses online shared cataloging by means of an entry form; and
10. A link exists to MetaLib, established in cooperation with Ex Libris (Uniform Information Gateway) via the HTTP protocol by means of the XML format.

Several problems remain at the time of writing:

1. The UC records are to be linked to the authority records of the cooperative authority database under construction in the Czech National Library;
2. The Z39.50 communication protocol is to be implemented; and
3. The Directory is to be transferred to CUBUS.

### Technological Parameters of the CUBUS System<sup>3</sup>

From the very beginning, it was clear that, once developed, CUBUS was going to have a single installation and would accommodate only the Union Catalog. This is why its development was entrusted to a smaller software company, whose programmers were better motivated to deal with the client's requirements. This option also brought about a reduction in the development costs. On the negative side, this choice involved higher risks related to the long-term stability of such a company. To minimize this risk, we adopted the following important steps:

1. To minimize dependence on a particular implementation team, we used only widely available technology (Oracle, PL/SQL, Java, HTML, XML); and
2. Professional software was created to document in detail the functions and structure of the system, which allows for rapid acclimatization of new staff (analysts and programmers) into the team.

### The Openness of the CUBUS System

Openness is understood as the ability to provide the necessary interfaces to interlinked systems. Such interfaces should accommodate the established de facto standards. From the very beginning, CUBUS was designed to focus on established de facto standards (TCP/IP, HTTP, FTP etc.), which has made it possible for the system to expand rapidly and widely. This also eliminated the high risks involved in using de jure standards (i.e. standards defined by a relevant commission) that had not been previously verified in real implementations.

---

<sup>3</sup> The following description was provided by an independent analytical company [1].

## The Flexibility of the CUBUS System

A system's flexibility consists in its ability to absorb modifications aimed at changing or extending its functions. A system's flexibility depends, first, on the technology used and, second, on the way it has been implemented (e.g. a system's modularity increases its flexibility). To achieve a sufficient level of flexibility, it is also important to provide consistent and easy-to-maintain documentation.

CUBUS provides sufficient flexibility because it employs widely available technology. This applies both to the database itself, built in Oracle, and to other tools used such as PL/SQL, Java, and servlets. Hence the system is not tied to a narrowly specialized development team, since the technology utilized is widely used and known. Considerable effort was also focused on a coherent conceptual approach to functionality requirements, which now makes it possible to easily expand the functions of the system. The system is continuously monitored by special software, a tool that allows quick and thorough analyses of the potential impact of any planned changes.

## The Scalability of the CUBUS System

Scalability is interpreted as the ability to increase a system's performance without having to modify it, or, in other words, the possibility to boost performance through a mere hardware upgrade and administrative operations. CUBUS' scalability is chiefly assured through using a robust relational system for database administration, Oracle, which is capable of absorbing several times more data than it currently holds without affecting the current speed of responses to users' queries.

During its development, CUBUS has not encountered any limitations. Its basic development has been completed both conceptually and practically. It is a system based on modern and widespread technological solutions (Oracle, Java, and servlets), and any improvements and modifications for changing users' needs are easy to carry out. In 2001, we had purchased a professional tool that made it possible to create a detailed data and process model of CUBUS, thus obtaining a high-quality description that is easy to understand. Since 2001, a full-scale copy of CUBUS is operated on a separate server that may be used as a backup when implementing an upgrade.

There is no serious reason for abandoning the system and having to face again the obvious limitations that ALEPH imposes when it is used for administering and operating a real-life heterogeneous union catalog.

### CUBUS as free software

The presentation of our paper at the Tallinn conference inspired one of the guest lecturers, Stefan Grandmann, to inquire about the ownership of the CUBUS system and whether there was any possibility of making the CUBUS system available as free software under the GNU GPL (General Public Licence) provision.

The author of the application, who owns the moral rights, and the National Library of the Czech Republic, which holds the economic rights, share the copyright of the CUBUS system. If the author of the application gives his written authorization, the National Library of the Czech Republic can consider making the CUBUS software available as free software under GNU GPL<sup>4</sup>.

### ALEPH + CUBUS (2002–?)

In April 2002, without identifying any factors that would mark the CUBUS system as an inadequate application, the Czech National Library decided to reverse course and operate the Union Catalog under ALEPH again, with the proviso that some modules (data import with duplicates testing, and online

---

<sup>4</sup> Software licences are mostly designed to take away the right to share and change the program freely. By contrast, the GNU General Public Licence is intended to guarantee freedom to share and change “freeware.” This GNU General Public Licence applies to most of the Free Software Foundation's software. A short quotation from the GNU GPL preamble: “When we speak of free software, we are referring to freedom, not price. Our General Public Licences are designed to make sure that you have the freedom to distribute copies of free software (and charge for this service if you wish), that you receive source code or can get it if you want it, that you can change the software or use pieces of it in new free programs; and that you know you can do these things.”

updating of series records) would nevertheless keep running under CUBUS, because otherwise they would have to be terminated.

The following key reasons were given for this decision:

- The Z39.50 protocol has not been implemented in CUBUS;
- There is no link from CUBUS to the authority files;
- It is not practicable for the Czech National Library to operate two systems (ALEPH and CUBUS); and
- CUBUS was produced by a small software company.

As for the first two issues, suffice it to say that both features should have been provided by the end of 2002 (see below for more detail) and had been planned in the projects submitted to the Ministry of Culture where we had applied for funding.

The third reason was dropped as soon as it had become obvious during a discussion at an Expert Council meeting that CUBUS cannot be entirely abandoned after all.

As to the fourth reason, it is evident that the small size of a company raises concerns about its long-term stability. But that concern needs to be contrasted by the fact that during the period in which the Union Catalog was operated under ALEPH, it proved impossible to induce Ex Libris, definitely a major and stable company, to implement any of the requested improvements to the Catalog. The reason is evident: an international software company which has several large installations in the United States, among others, is unable to deal with a single requirement of a minor customer somewhere in the heart of Europe. This may make one ponder what is more beneficial for the Union Catalog in the Czech Republic: to be an insignificant minor customer of a major software company, or to be a major customer of a minor but stable software company?

### How Will the Union Catalog Function under the ALEPH-CUBUS System?

The transfer to ALEPH will take place in two stages:

In the first stage (expected completion by January, 2003), CUBUS will process the members' batch imports during the daytime, while at night it

will export the new or updated records in the RUX format.<sup>5</sup> The file will be imported into the ALEPH database, the records of which will be updated according to the system number. Users will be able to search and export records under the ALEPH system.

In the second stage of the CUBUS–ALEPH transition, after installing the Z39.50 protocol containing the Update function on ALEPH,<sup>6</sup> only those libraries which have implemented the Z39.50 protocol will be able to catalog their monograph and special document records online in the Union Catalog. Every day at 7 pm, ALEPH will close the data editing function in the Union Catalog and will export the new or newly edited records. This file will be imported into CUBUS by a standard method, thus assuring the congruence of the two databases. Then CUBUS will turn to the batch record imports that arrived during the day. By 4 am, the newly imported records will be exported in RUX format and they will update the ALEPH database according to their system number.

The serials records will continue being updated online in CUBUS without having to close the database to users [11].

The Union Catalog administrator was not invited to attend any of the policy discussions of the Czech National Library's top management about the Union Catalog platform change. Instead, she made use of a Union Catalog presentation to inform the management about the potential problems of operating a Union Catalog under ALEPH in combination with CUBUS.

A functional connection between the two systems would present the Union Catalog administrator with the following problems [25]: the National Library's need for programmers will increase; the number of ALEPH licenses needed will have to grow<sup>7</sup>; hardware requirements for ALEPH will grow in excess of planned system upgrades; data administration (especially

---

<sup>5</sup> RUX is an internal format for data import into ALEPH and is based on the structure of textrow UNIMARC in UTF-8 + RecID.

<sup>6</sup> Although it appears that Ex Libris has totally abandoned the development of a Z39.50 protocol containing the Update function.

<sup>7</sup> CUBUS is licensed for an unlimited number of access points.

imports) will continue to be provided under CUBUS; and the processing of series under different library systems will be highly problematic. Hence, series processing including their online updates is to remain under CUBUS, which will still have to be maintained and developed. Finally, problems that exist in local library systems may compromise the functionality of the national Union Catalog and vice versa.

Users will not be offered new functionalities or benefits; on the contrary, the change will cause service to deteriorate in many respects. Thus, the Union Catalog's handling of series, ILS, and the Directory is bound to suffer, and the existing unlimited license for access with no time restrictions will be replaced by one with time restrictions. Furthermore, the benefit derived from the existing connection with the Universal Information Gateway (UIG) through HTTP and XML will be lost.<sup>8</sup> CUBUS-to-ALEPH data imports will have to be carried out overnight, for which the National Library's databases may be closed to users, although statistics show that the Union Catalog is accessed at night as well (e.g. by users from distant time zones, or by those waiting for a freer and faster Web). Under CUBUS, imports are conducted in the background, and so there is no need to close the database. With ALEPH, users will be provided only with an out-of-date copy of the Union Catalog, which is not only a non-standard approach, but is counterproductive and to be used only for very serious objective reasons.

The reasons that made the Czech National Library start to develop its own Union Catalog system in 1997 still exist. Even in combination with CUBUS, ALEPH does not have the tools to successfully operate a heterogeneous Union Catalog. ALEPH is a very good system designed to administer and operate a library, but not suited for a real-life heterogeneous union catalog.

The Union Catalog run under ALEPH will again be limited by the capabilities of ALEPH, which—although broader than in 1997—are still unable to react flexibly to the requirements and demands of its customers, in contrast to the functionality of CUBUS.

---

<sup>8</sup> The existing UIG connection does not impose license requirements on the use of the Z39.50 protocol within UIG.

## 9 The Connection between Authority and Bibliographic Records in the Union Catalog

The National Authority Department at the Czech National Library accomplishes two basic objectives:

1. The administration of local authorities at the Czech National Library; and
2. The formation of a national authority database for libraries of the whole Czech Republic.

In December 2000, representatives of Czech libraries and library system vendors met at the Czech National Library to discuss the cooperative creation of a national authority database. This discussion ended with a clear and unambiguous recommendation to build the national authority files within the Union Catalog system, i.e. under CUBUS, after which the Union Catalog administrator drafted a document that identified the fundamental policies in creating a cooperative national authority file. This document was posted on the CASLIN website by the end of 2000 [14].

Simultaneously, the administrator presented a proposal to the management of the Czech National Library for a reorganization [20] that included the transfer of activities related to the creation of a national authority database to the administration of the Union Catalog. This proposal was rejected.

In January 2001, the top management of the Czech National Library announced that the authority files would be built under ALEPH [5]. This denied the opportunity to build a single system (CUBUS) for both bibliographic and authority records within the framework of the Union Catalog. Subsequently, alternative paths were explored.

At first, the viability of linking the bibliographic records from the CUBUS-based Union Catalog with the ALEPH-based authority database was examined. It turned out that even ALEPH 500 does not permit record modifications using tools other than its own.<sup>9</sup> Hence, it was impossible to

---

<sup>9</sup> Such modifications have to be carried out outside the ALEPH database, as was the case with ALEPH 300.

set up an online link between the bibliographic records from the CUBUS-based Union Catalog with the ALEPH-based authority database. The Union Catalog administrator proposed and provided a detailed description of a solution whereby a copy of the authority database would be defined within the CUBUS system. The database copy would be updated every night, so as to achieve congruence with the ALEPH master [15]. This would have enabled CUBUS to establish a link between the bibliographic records in the Union Catalog and the authority records, thus making the records in the Union Catalog, and the authority database—all very straightforward authority records—available to all Union Catalog users regardless of which system they were using. The Union Catalog administrator's objective was to expand CUBUS with functionalities for gathering and administering authority records, closely linked to functionalities for work with bibliographic records. By storing the bibliographic and the authority records in one database, the overall complexity of the system would be reduced, since measures to treat duplicates and inconsistencies created during data transfer between different systems would become unnecessary. It would also provide an opportunity for establishing a link between the data. The increase of complexity under the alternative solution also increases the probability of errors, and hence of costs. The biggest benefit of storing the bibliographic and the authority records in one database would stem from the reduction of processing costs. It would be possible to make use of the existing applications for bibliographic record administration, which would also allow us to create a two-way link between the bibliographic and the authority data. Such a link would make it possible to display all relevant bibliographic records for a given authority.

## 10 The Union Catalog and the Z39.50 Protocol

As late as 2000, there were extremely few libraries in the Czech Republic that had a truly functional Z39.50 protocol implemented in their systems [24]. Basically, only the ALEPH libraries, probably four in number, had it, which is why the implementation of Z39.50 protocol was low on the list of priorities during the development of CUBUS. At the same time, the Union Catalog administrator was aware of the fact that due to the inherent

openness of CUBUS, the Z39.50 protocol would not provide the system with any new functionality and would not help other entities to join the Union Catalog and enhance its expansion. The pressure on the part of the Czech National Library's management to implement the Z39.50 protocol was not, at the time, based on any existing needs of the Union Catalog members.

In 2001 and 2002, rather than developing its own Z39.50 protocol, the Union Catalog administrator was planning to implement one designed by an independent software company. For successful implementation of a protocol from a different company, it was crucial to describe the existing CUBUS system, and so the Union Catalog Department staff created a document called "A Functional Model of the CASLIN Union Catalog" [15].<sup>10</sup> Although the model was correct and effectively described the state of affairs and anticipated CUBUS developments, it did not explicitly differentiate the process and data models of the system, and the data flow diagrams were not based on standard description tools (which were not available to the Union Catalog administrator at the time), which is why this functional model was not sufficient for the Z39.50 vendor. The only logical solution was to entrust the creation of CUBUS' data and process models to an outside group of consultants. The group completed the analysis of the current state of CUBUS, including a forecast of its future connection with the authority database and of the implementation of the Z39.50 protocol. The group also prepared a documentation of CUBUS, which included a rational approach to recording changes. This tool is still available and at the Union Catalog administrator's disposal.

## 11 The Union Catalog in the Uniform Information Gateway

MetaLib, the Uniform Information Gateway (UIG) software by Ex Libris, may be connected to sources through the Z39.50 protocol or through the HTTP protocol. To set up a connection through the HTTP protocol is more

---

<sup>10</sup> Available at the Library and Information Sciences' Reading Room at the Czech National Library.

difficult, and so Ex Libris is usually willing to provide it only for world-class information sources. Connecting local sources accessible only through a HTTP protocol is quite expensive, because it requires more programming on the part of Ex Libris [18]. The Union Catalog belongs to the latter class, and we are grateful to Ex Libris for connecting it without additional charges. The Czech Libraries appreciate the presence of a national information resource in the UIG, and UIG, in turn, adopted the CASLIN logo, which is well-known not only in the Czech and Slovak Republics, but elsewhere as well.

The Union Catalog was connected to UIG through an external program assuring the conversion of a query and its result between the two systems [26].

A query placed by a UIG user is sent to the Union Catalog through the HTTP protocol<sup>11</sup> coded in UTF-8 and in the format `http://server address?access file=value&access file=value`. The output is in the form of a record exported in XML. The extent of the record is a compromise between the data displayed in MetaLib and the specifics of the Union Catalog in question, i.e. the record contains the basic identification data of the document plus the name of the library that owns it, and for periodicals also the year.

For the output in XML, the Union Catalog administrator created a DTD<sup>12</sup> that complies with the current Union Catalog requirements for a connection to UIG. In the future, it will pose no problem to expand the proposed DTD or modify it in order to comply with the worldwide accepted definition of a document type for library formats.

The method of establishing a connection to UIG via HTTP and XML will be also used to create a connection to other sources that neither work with the ALEPH system nor use the Z39.50 protocol.

---

<sup>11</sup> See <http://www.caslin.cz:7777/caslin/ENG/parameters.html>.

<sup>12</sup> DOCTYPE CaslinMeta. For more information see <http://www.caslin.cz:7777/caslin/ENG/dtd.html>.

## A Minor Excursion into Real and Virtual Union Catalogs

The introduction of computers, library systems, and the MARC format represented a quantum leap in the accessibility of information in union catalogs in comparison with their card-based predecessors. And while the tasks and problems in administering a union catalog<sup>13</sup> have not changed, the cost of administering and operating electronic union catalogs has probably changed for the worse. The high cost of operating a real union catalog and the proliferation of the World Wide Web are probably the main reasons for the appearance in the early 1990s of the first information gateways. In the mid-1990s, two nationwide virtual union catalogs emerged in the Czech Republic:

1. The homogeneous TinWEB, established as a system for parallel searching of library catalogs within the TINLIB system; and
2. The heterogeneous ATpar (ALEPH-TINLIB Parallel Query System), designed and implemented for single-query transparent searching of a selection of library catalogs with an ALEPH-based and TINLIB-based WWW interface [2].

In the library community, access points to Z39.50-based sources are established via information gateways based on WWW, where the end-user interface is a browser (in 2001 in the Czech Republic, it was UIG).

In general, an information gateway may be considered a virtual Union Catalog only when it complies with the following requirements (from the end-user perspective):

- Access takes place through a single user interface;
- The format for search queries uses a unified syntax, using a single set of search attributes;
- The records are provided in a single, common format allowing a single record to be displayed;
- Duplicate records are not displayed;

---

<sup>13</sup> For example data import, duplicates testing, accounting for system heterogeneity.

- All records are available with location data and the full set of data about the library collection;
- All copies are provided with up-to-date loan status and availability data; and
- A virtual union catalog is interlinked with several integrated loan and order services [8].

Ordinary users trying to identify and locate a document particularly appreciate information gateways, functioning as virtual union catalogs. For inter-library loan services and shared cataloging, professional librarians prefer real union catalogs with records from tens or hundreds of libraries that have been classified and evaluated and made available in one database.

It should be noted that real union catalogs provide their users with feedback in the form of error messages, i.e. information about deficiencies in delivered records. Thus the participants have the possibility to improve the quality of document processing in their home institutions. At the same time real union catalogs support the implementation of uniform standards. Virtual union catalogs use a technology that moderates and compensates for the differences among the various participants' systems. Virtual union catalogs do not provide any feedback, its participants do not have the need to improve the quality of its records and there is no need to implement uniform standards.

## 12 Union Catalog Terms of Payment

Both access to and use of the Union Catalog are free of charge. The members do not obtain any fees for the records they provide, and conversely, they do not pay anything for importing Union Catalog records, which they use for cataloging or retroconversion of their collections [22].

The Czech National Library funds the administration and operation of the Union Catalog, i.e. the Union Catalog is fully state-funded through the Ministry of Culture of the Czech Republic.

## 13 Union Catalog Organization and Control

### Union Catalog Administration in the Czech National Library

In connection with the launch of CASLIN and based on the assumption that the CEZL Department (National Registry of Foreign Literature) would become the administrator of the CASLIN Union Catalog, CEZL staff articulated a detailed “Strategy for the CEZL Union Catalogs’ Transformation into the CASLIN Union Catalog” [4]. Based on this strategy, on July 1, 1994, the CEZL Department became, for a short but very useful period of 11 months, a part of the Cataloging Division, where CEZL staff learned the details of the newly discussed standards. As early as June 1, 1995, the CEZL Department became an independent Union Catalog Section (with CEZL and CASLIN as the union catalogs to work with) and it became directly answerable to the Director of the Czech National Library. Currently, the Deputy Director of the Library heads the independent Union Catalog Division [16].

### Union Catalog Task Teams

As a part of CASLIN, the CASLIN Task Team for the Union Catalog was formed jointly with Slovak librarians in January 1994. This task force produced high-quality strategic documents dealing with union catalog administration, architecture, construction, and member typology. Due to the ever-increasing scope of its activities, the Task Team’s decision-making became too cumbersome, and so in November 1995, the Union Catalog Task Team was established, this time without Slovak participation. This Team has been in operation to the present day.

The cooperation between the Union Catalog and libraries is strictly regulated on a contractual basis. Since October 1996, the agreement concluded between the Czech National Library and the cooperating libraries has been called “Cooperation Agreement on CASLIN Union

Catalog.”<sup>14</sup> Consulting bodies available to the Union Catalog administrator are

1. The Union Catalog Task Team, with staff consisting of representatives from libraries that supply the Union Catalog with records, since November 1995;
2. The Union Catalog Research and Development Team, and expert body established in June 1997 by the Union Catalog administrator, whose main task was to overcome technical difficulties arising in the Union Catalog. In late 1997, the R&D Team discussed the issues related the development of a new tailor-made Union Catalog system;
3. The Union Catalog Expert Council, formed in December 2000, which absorbed most of the R&D Team’s members.

The members of the Union Catalog Expert Council played a decisive role in formulating the changes in the Union Catalog system. The work of the Council has been adversely affected by a variety of conflicts of interest—to wit, the ALEPH sales representative for the Czech Republic is also a member.

Despite the diversity of opinions and interests, the first vote was in favor of CUBUS [6], and the second ended in a 6:6 draw [7]. It is also the case that that vote was characterized by certain irregularities; for example, the absentee vote of a member was counted, the ALEPH sales representative’s vote was not disqualified, etc. The Union Catalog administrator participated in the Council only as a non-voting secretary. This was the vote on the basis of which the management of the Czech National Library decided to replace the current platform of the Union Catalog, although the Union Catalog administrator expressed her disagreement with the decision.

## 14 The CASLIN Consortium

The cooperation underlying the CASLIN Union Catalog is legally based on the Cooperation Agreement on CASLIN Union Catalog. During 2000, the

---

<sup>14</sup> Its text is available on the CASLIN website.

Union Catalog Administrator undertook specific efforts to create a CASLIN Consortium.<sup>15</sup> The item on the Association's agenda was to be the Union Catalog.

However, conditions in the Slovak Republic did not permit the creation of a functional international CASLIN Consortium based on union catalog cooperation. The very name of the consortium proved to be a contentious issue, since its acronym contained the letter "S" standing for Slovak, although none of its would-be members was Slovak. Although imprecise, the existing acronym acquired such a familiar status in the Czech library environment that it seemed counterproductive to abandon it. In this situation, during the preliminary discussions, the Union Catalog members themselves suggested a new name that would fit the original acronym: CASLIN would stand for the Czech Association for Services in Library Information Network.

Developments on the Slovak side seem to indicate that the their union catalog initiatives have been following a different path from their Czech CASLIN counterpart, both in relation to the speed of development and to its strategy. Nevertheless, the crucial orientation established at the beginning of the joint CASLIN Project in 1993 has been preserved. There is always a real possibility that the CASLIN Consortium will expand and become, once again, a functioning international body, thus returning to its roots and forming a truly Czech and Slovak library information network. The idea of founding a union catalog-based CASLIN Consortium has been abandoned, since it is now part of the legal mandate of the Czech National Library (the new legislation was passed in 2001) to build and run a national Union Catalog.

## 15 Conclusion

There are no signs that suggest that, on a worldwide scale, real union catalogs are beginning to fall into disuse, not even in countries where

---

<sup>15</sup> A draft of the Statutes of the CASLIN Association was drawn up, and the Cooperation Agreement was amended.

information technology is very advanced and virtual union catalogs are relatively easy to build. Real union catalogs represent a unique tool for value-added services expected and required by both users and librarians. Virtual union catalogs provide additional functions (especially those of location) suitably complementing their real counterparts.

The Czech National Library has not set out to build its virtual union catalog (i.e. the Uniform Information Gateway) with the aim of providing a supplementary service to the existing real Union Catalog. It is meant to become the primary union catalog format, which is corroborated by the decision to stop developing CUBUS as a real union catalog system. This approach has de facto compromised the ability of the union catalog to provide equal service to all users and, above all, to libraries with different library systems.

## References

- [1] *Analýza systému CUBUS*. LBMS, c2002, p. 16.
- [2] Bartošek, Miroslav. "ATpar: Aleph-Tinlib Heterogeneous Virtual WWW Catalog." In *Union Catalogs: their organization and services. Zadov (Czech Republic), 30th–3rd 1999*. Praha: Národní knihovna ČR, 2000: 47–53. ISBN 80-7050-354-8. Also at <http://www.caslin.cz:7777/caslin99/s2.htm>.
- [3] Dvořáková, Helena. "Budování souborného katalogu s ohledem na možnosti dané programem na řízený import s kontrolou duplicit." Unpublished paper. Praha: Národní knihovna ČR, 13.11.1996.
- [4] Dvořáková, Helena, Gabriela Krčmařová, and Lydia Vítová. "Koncepce transformace CEZL v souvislosti se soubornými katalogy CASLIN." Unpublished material. Praha: Národní knihovna ČR, 21.3.1994.
- [5] *Expertní rada pro Souborný katalog ČR: zápis I.jednání dne 24.1. 2001*. Praha: Národní knihovna ČR, 2.2. 2001. Available at <http://www.caslin.cz:7777/caslin/expert1.html>.

- [6] *Expertní rada pro Souborný katalog ČR: zápis 2. jednání dne 27.2. 2002*. Praha: Národní knihovna ČR, 11.3.2002. Available at <http://www.caslin.cz:7777/caslin/exp0202.html>.
- [7] *Expertní rada pro Souborný katalog ČR: zápis 3. jednání dne 20.3.2002*. Praha: Národní knihovna ČR, 26.3.2002. Available at <http://www.caslin.cz:7777/caslin/exp2003.html>.
- [8] Hajná, Kateřina. "Testy na duplicitu klíčů v souborném katalogu." Unpublished paper. Praha: Národní knihovna ČR, březen 2002:4.
- [9] Hajná, Kateřina et al. "Časová a finanční rozvaha vývoje systému CUBUS v roce 2002 (2003)." Unpublished paper. Praha: Národní knihovna ČR, březen 2002.
- [10] Hajná, Kateřina et al. *Zadání pro vývoj aplikací v systému ORACLE: tištěné seriály*. Praha: Národní knihovna ČR. Available at <http://www.caslin.cz:7777/caslin/caslin-ser.html>.
- [11] Hajná, Kateřina and Gabriela Krčmařová. "Záměr rozvoje CASLIN—Souborného katalogu ČR v souvislosti s přechodem na provoz v systému ORACLE 7: materiál na poradě vedení NK ČR dne 24.9.1997." Unpublished paper. Praha: Národní knihovna ČR, 5.9.1997:5.
- [12] Hajná, Kateřina and Gabriela Krčmařová and Danuše Vyorálková. *Zadání pro vývoj aplikací v systému ORACLE: tištěné monografie*. Praha: Národní knihovna ČR, 1998. Available at <http://www.caslin.cz:7777/caslin/zadani2.html>.
- [13] Hajná, Kateřina and Eva Svobodová. *Budování báze autorit v prostředí CASLIN—Souborného katalogu ČR (SK ČR)*. Praha: Národní knihovna ČR. Available at <http://www.caslin.cz:7777/caslin/autor-zad.html>.
- [14] Hajná, Kateřina and Eva Svobodová. *CASLIN—Souborný katalog ČR: funkční model*. Praha: Národní knihovna ČR, únor 2001: 36.
- [15] Husby, Ole. "Real and Virtual Union Catalogs." In *Union Catalogs: their organization and services, Zadov (Czech Republic), 30th–3rd 1999*. Praha: Národní knihovna ČR, 2000: 112–120.

- ISBN 80-7050-354-8. Also available in HTML version at <http://www.caslin.cz:7777/caslin99/a2.htm>.
- [16] Krčmařová, Gabriela. "35 let historie CEZL a současnost v rámci CASLIN." *IT lib: informačné technológie a knižnice*, 4 (2000). ISSN 1336-0779. Available at [http://www.cvtisr.sk/itlib/bc2000\\_4/krcmarova.htm](http://www.cvtisr.sk/itlib/bc2000_4/krcmarova.htm).
- [17] Krčmařová, Gabriela. "CASLIN—The Union Catalog of the Czech Republic," In *Union Catalogs: their organization and services, Zadov (Czech republic), 30<sup>th</sup>–3rd 1999*. Praha: Národní knihovna ČR, 2000: 158–164. ISBN 80-7050-354-8. Also available in HTML version at <http://www.caslin.cz:7777/caslin99/a7.html>.
- [18] Krčmařová, Gabriela. "CASLIN—Souborný katalog ČR: materiál na poradu vedení dne 21.11.2001." Unpublished paper. Praha: Národní knihovna ČR, 19.11.2001:3.
- [19] Krčmařová, Gabriela. "Kam kráčí CASLIN- Souborný katalog ČR." In *Automatizace knihovnických procesů—VII, sborník ze 7. ročníku semináře pořádaného ve dnech 9.-10. června 1999 v Ústí nad Labem*: 59–69. ISBN 80-238-4233-1.
- [20] Krčmařová, Gabriela. "Návrh na organizační změnu oddělení pro souborné katalogy." Unpublished material. Praha: Národní knihovna ČR, 18.12. 2000+:2.
- [21] Krčmařová, Gabriela. "Od systému ALEPH k systému ORACLE." In *INFOS'98. 2001: zborník z 28. medzinárodného infromatického sympózia, ktoré se konalo v dňoch 27.-29. apríla 1998 v Starej Lesnej*: 66-73. ISBN 80-85165-67-8. Also available in HTML version from <http://www.caslin.cz:7777/caslin/wwworac.html>
- [22] Krčmařová, Gabriela. "Rozvaha a návrh financování CASLIN— Souborného katalogu ČR." Unpublished paper. Praha: Národní knihovna ČR, srpen 1998:6.
- [23] Krčmařová, Gabriela. "Souborné katalogy v heterogenním prostředí." In *INFOS 2001: zborník z 31. medzinárodného infromatického sympózia*,

*ktoré se konalo v dňoch 2.–5. apríla 2001 v Starej Lesnej:* 310-318. ISBN 80-85165-84-8.

- [24] Krčmařová, Gabriela. “Z39.50—protokol z minulého století.” *IT lib: informačné technológie a knižnice*, (4) 2001. ISSN 1335-793X. Available at <http://www.cvtisr.sk/itlib/itlib014/z3950.html>.
- [25] Krčmařová, Gabriela and Eva Svobodová. “Prezentace CASLIN—Souborného katalogu ČR na PV dne 5.4.2002.” Unpublished paper. Praha: Národní knihovna ČR, 2002:3.
- [26] Krčmařová, Gabriela and Ilona Trtíková. “Funkcionalita CASLIN—Souborného katalogu ČR jako centralizovaného zdroje pro informační profesionály a katalogizátory.” In *INFORUM 2002: 8. konference o profesionálních informačních zdrojích, 21.-23. května 2002, Nová aula Vysoké školy ekonomické, Praha*. Available at <http://www.inforum.cz/inforum2002/prednaska15.htm>.
- [27] *Zápis 3. jednání Pracovní skupiny pro souborný katalog při NK ČR*. Praha: Národní knihovna ČR, 3.6.1997.

## Links

<http://www.caslin.cz>—Website of CASLIN Union Catalog.

<http://psi.nkp.cz:2400/r/SKK/p210/pcz>—Access to CUBUS:

CASLIN Union Catalog—Monographs and special types of documents.

CASLIN Union Catalog—Serials.

<http://sigma.nkp.cz:4525/ALEPH0/~~/START/adr>—Access to Aleph:

CASLIN Union Catalog—Directory of libraries and information centers of the Czech Republic.

<http://www.caslin.cz:7777/caslin/historie/document.html>—Documents about CASLIN.

<http://www.caslin.cz:7777/caslin/dtd.html>—Description of CaslinMeta.

<http://www.caslin.cz:7777/caslin/parameters.html>—Description of query CUBUS.

[http://jib-info.cuni.cz/dokumenty/dokumenty\\_tech.html](http://jib-info.cuni.cz/dokumenty/dokumenty_tech.html)—Specification of http interface for connecting to Metalib.

<http://dior.ics.muni.cz/hales/atpar/>—ATpar (ALEPH-TINLIB Parallel Query System).

<http://sd.ruk.cuni.cz/tinweb/sd/k6>—TinWEB.

## **Chapter 8**

# **LINCA: The Union Catalog of the Czech Academy of Sciences**

Martin Lhoták

The Academy of Sciences of the Czech Republic (ASCR) is the largest non-university scientific institution in the Czech Republic. It comprises the Main Library and the libraries of 60 basic research institutes.<sup>1</sup> Almost every Institute of the ASCR has its own topically focused library; some Institutes even have more than one library. These libraries are an indispensable resource of information for scientists. The ASCR also maintains a central library totaling 1,000,000 books and periodicals.

In 1992, the main library and a large number of institute libraries accepted a UNESCO grant offered to libraries in the Czech Republic to install the library system Micro-CDS/ISIS, which was a good solution at that time. Micro-CDS/ISIS had low hardware requirements and the system enabled broad customization. It was possible to run the system on local PCs or in the local network.

Large investments of grant moneys permitted the completion of a high-speed connection to the Internet for almost all Academy institutes in 1996. It opened new opportunities and new services for librarians, information workers and library users.

In 1996, the Library obtained support from The Andrew W. Mellon Foundation for the LINCA project, the Library Information Network of the Czech Academy (of Sciences). This was a turning-point because it provided a great opportunity for building the Union Catalog (UC) of the ASCR—a

---

<sup>1</sup> See Table 1 in the Appendix.

catalog that anyone could search in each of the Academy's 65 libraries from a single location. The LINCA project has as its goal the construction of a library information network based on new hardware and software, and the creation of the UC. There were not many different library software systems being used in the Czech Republic at that time. ALEPH or TINLIB were the most frequently installed ones in large Czech libraries.

BIBIS was the newest library system on the Czech market. Although we initially thought about the ALEPH system, there was a problem with ALEPH's local branch, and this fact assigned a lower evaluation to that system in the selection process. All three systems had similar levels of quality and user interface. The final choice was BIBIS. The system had a good functionality, was capable of broad customization and also had good references from abroad (e.g. Philips). BIBIS also had a local distributor, INCAD Ltd., in Prague. These were the main reasons for ASCR's choice. Connection to the server was to be via telnet for BIBIS. In 1996, this was a reasonable solution, considering that some Institutes had a slow Internet connection and low-performance end-user stations. Now, however, it appears to be a serious limitation. The library could choose either a central or a partly distributed system. The distributed system was chosen, based on 14 servers from Sun Microsystems. The project implementation group figured that this choice would offer more room for customizing the final setup.

An additional reason for the choice was that connections between some of the Institutes were still not fast enough. This solution did not seem to be ideal in some respects. The main problem was to find qualified administrators for some of the servers. The new system was at first planned only for Academy institutes that were interested and wanted to participate in the project. But the Academic Council of ASCR decided to fund the new information system to include all ASCR institutes. It was a good idea, but individual ASCR institutes now have a relatively high level of administrative independence, and it was very difficult to initiate and maintain cooperation and communication when some of them indicated no interest in participating. But gradually, more institutes began to work with the system. Its success was (and still is) very dependent on the individual librarians' efforts and abilities at each library. Considerable problems emerged when the Micro-CDS/ISIS system was in operation and when it was necessary to convert

data. The data structure was often not standard, because customized Micro-CDS/ISIS installations varied among individual Institute libraries, and there was also a lot of duplication in authorities and thesauri. It was also hard to find enough Micro-CDS/ISIS specialists who could ensure high-quality data export from the system.

Considering the capabilities that BIBIS had to offer, it was decided in 1997 to create a 'physical' union catalog. The advanced search system Excalibur Retrieval Ware (renamed Convera Retrieval Ware in 2001) was chosen for the UC OPAC tool. BIBIS' vendor in Prague also distributes and services this system, which facilitates making a direct data export from BIBIS to the Union Catalog. The data from the Micro-CDS/ISIS system can also be converted to the UC and it is possible to continually update them, but some data corrections need to be made. We started with 10 cooperating institutes in 1997. In 2002, the UC contained 250,000 entries from 40 institute library databases. In addition, card catalogs were scanned, and the scanned catalog contains more than 1,000,000 cards. The UC makes it possible to order publications through ILL from the main library. In the future we hope to have all 65 libraries in the UC and offer more services. But BIBIS in its Version 97 is quite limited in what it can offer. For example, it is not possible to provide certain information or services to users, such as whether a certain book is physically in a library or to reserve a book. These features are not typically available through the UC, but would, of course, be appreciated by our readers. In institutions having a central administration such as the ASCR, it might be possible to accomplish this. It could be done through strong cooperation and by centralizing management in one place—say, the main library. In time we would like to integrate all ASCR Institute libraries, but it depends on their general willingness to cooperate. Another problem in some institute libraries was the lack of professional librarians with sufficient technical background. It was noted at the beginning of the LINCA project that finding and recruiting enough trained, experienced, quality people is much harder than finding funding for the project. This has been fully confirmed over time.

A technical solution other than BIBIS must be found in the near future, because we are unable to offer advanced services (online reservations, Z 39.50, etc.) that are in fact the standard in the latest systems. The way to go

is to build a central system. This would ensure a more efficient operation for the entire system, for the Union Catalog and also for each institute library database.

## 1 A Closer Look

BIBIS has been adopted by approximately 20 libraries in the five years since its first installation. Forty of the 65 libraries are contributing to the UC of ASCR at the present time. While our ambition for expanded use of BIBIS at institute libraries was originally much higher, it is important to recognize the number of institutes that are now part of the system. Furthermore, one third of the institute libraries are using a system that obliges the user-librarian to observe all necessary cataloging rules, which is another significant achievement. A majority of our BIBIS users have not had any system before and have cataloged only to paper cards.

Of the several problems that remain, data conversion, the implementation of a distributed system, and human resource management are key. All of these are aggravated by financial problems.

## 2 Data Conversion

The Main Library and most institute libraries had cataloged in Micro-CDS/ISIS before the advent of BIBIS. A merit of Micro-CDS/ISIS is its flexibility. However, the problem was the transition to a new system, because no single procedure of conversion was appropriate for all libraries. Furthermore, libraries often did not have their own supervisors. There was only one person in the entire ASCR able to administer ISIS installations and who was qualified to provide system services. It was not possible to convert to BIBIS in all institutes, because many had no staff member to provide technical supervision. As a result, most of the libraries just gave up.

### 3 Distributed System

An improvement in the Academy's computer network occurred during 1995–1997. This improvement made possible a solution in which several servers would be connected to neighboring libraries. However, the bandwidth of the network at that time did not permit the use of a single central server. Administrators of servers were assigned during the planning and organizational stage; one for technical aspects and one for dealing with librarians. The startup and service of individual installation depended on the quality, interest and willingness of each administrator. The LINCA staff present at the beginnings had high resolve and determination. However, nobody imagined that some might not share that enthusiasm and that administrators of these servers might not be adequately remunerated. Although the problem of remuneration was solved in some cases, everything tended to come to a halt at institutes where staff did not have a strong interest in the new system. Even in libraries with good server administration, there tended to be problems because of data conversion difficulties.

### 4 Human Resources

Plans for the deployment of human resources were often inadequate to insure the proper implementation of the project. Complications would arise during data conversion or with server administration when staff was not competent to handle the job. The librarians in certain libraries often failed to get high-quality administrators for their servers, which prevented them from starting the cataloging module. But as time went on, some of the Institutes began to work with BIBIS anyway. Unfortunately, some of them became discouraged and quit.

A final problem arose as a result of inadequate BIBIS training. Since there were not many qualified librarians at the ASCR, most were unable to catalog in conformity with the standards (AACR2, UNIMARC) that were mandatory for BIBIS. Since no training courses had been organized for this purpose, librarians had to do much more work than normal. They were able to participate in training courses organized from time to time at the

National Library, but the initiative for this had to come from the individual librarians.

## 5 Motivation

People with varying levels of motivation were joining the project. The implementation team was highly organized and exhibited a strong sense of responsibility. But some libraries joined the project without a strong interest in it. As is often the case, the novelty of the system induced anxiety in some librarians for whom the system represented a set of brand new library standards. The motivation of server administrators varied from case to case. Cooperation was not as timely or as close as might be desired, because in some cases administrators were not librarians, but outside employees. Aversion to changing traditional methods of work and to learning new skills caused significant disruptions in the Cataloging Department of ASCR's Main Library when Micro-CDS/ISIS was installed. The same disruptions took place again during the transition from ISIS to BIBIS.

## 6 Finances

ASCR succeeded in obtaining significant financial resources from various sources (the Mellon Foundation and the Grant Agency of the ASCR) for acquiring hardware and software. But additional funds that could have been used for solving the conversion problems at the various institute libraries and for improving server administration did not become available.

In spite of these problems, the Union Catalog was created. Catalogs of individual institute libraries have been imported into a central database. It is now possible to search almost two-thirds of ASCR Institute libraries. The time it takes for new records to appear depends on how quickly the institutes submit the data. Some libraries update their records every 2–3 weeks, others only annually. The Union Catalog currently permits simple and advanced searches. It does not provide other functionalities, such as downloading of records, document delivery service or book reservation. The results of searches are displayed in the ISBD format.

## 7 Need for Changes

During the five years that BIBIS has been in use, it has not upgraded. As a result, it is becoming outdated and is increasingly incapable of performing the services that library users expect today. A case in point is the Uniform Information Gateway project (UIG). This project enables uniform access to electronic information resources, including the catalogs of the large Czech libraries. UIG administrators require participating libraries to run their own Z39.50 server. If a library has not installed this server, UIG administrators cannot integrate the library's catalog into the gateway. Although BIBIS promised that Z39.50 would be installed in an upgrade, Z39.50 is still not operational. With the current version of BIBIS, it is impossible to ensure readers' access to the status of their account or to provide them with information about the accessibility of publications turned up in a search. Shared cataloging is possible in BIBIS, but it is limited to cooperation among the users of each individual server. By now, a new modern library system should handle all these features. Since BIBIS does not, ASCR has to select a new system. The installation of very fast Internet connections to all institutes makes it possible, for the first time, to contemplate a fully centralized system.

Such a solution has several advantages. Since the system would need only one or perhaps two servers, hardware costs would decrease. The number of hard-to-find administrators and personnel costs would also decrease, while average competence would increase. A centralized module would also permit full use of records introduced by other libraries in the system. By utilizing holdings information, many duplicate records could be eliminated and storage requirements on the server would also decrease.

Much benefit is likely to be derived from selecting a system that is already in use at the major research libraries of the Czech Republic. The level of satisfaction with the newest version of ALEPH is high, and barring financial problems, that is the direction in which the ASCR libraries need to move. Introduction of this system at the Main Library and some institute libraries could take place in 2003, and the Union Catalog and the remaining institute libraries could come online in the new system in 2004.

## Appendix

**Table 1. Institute Libraries**<sup>2</sup>

Server	Abbreviation	Name of Institute	Size of Library
<b>KNAV</b>			
	LINCA	Union Catalog of the Institute Libraries of the ASCR	3,500,000
	KNAV	Main Library of the ASCR	945,412
	NHU	Economics Institute	62,500
	USMH	Institute of Rock Structure and Mechanics	27,700
	UDU	Institute of Art History	66,300
	UE	Institute of Electrical Engineering	4,100
	UH	Institute of Hydrodynamics	15,500
	UJC	The Czech Language Institute	57,400
	USP	Institute of State and Law	39,300
	UCL	Institute of Czech Literature	116,500
	A	Archives of the ASCR	69,000
	SLU	Institute of Slavonic Studies	69,990
	USD	Institute for Contemporary History	16,900
	UOCHB	Institute of Organic Chemistry and Biochemistry	82,000

---

<sup>2</sup> Ivana Kadlecova, "Report from projects," Library of ASCR, Prague, 1998.

<b>Server</b>	<b>Abbreviation</b>	<b>Name of Institute</b>	<b>Size of Library</b>
<b>KNAV</b>	UMG	Institute of Molecular Genetics	see UOCHB
	UHV	Institute of Musicology	22,400
		Size of Database	1,595,002
<b>SLOVANKA</b>			
	FZU-S	Institute of Physics	27,870
	UFP	Institute of Plasma Physics	10,900
	FZU-C	Institute of Physics	65,630
		Size of Database	104,400
<b>ZITNA</b>			
	MU	Mathematical Institute	66,840
		Size of Database	66,840
<b>JILSKA</b>			
	CTS	Center for Theoretical Study	20,000
	FLU	Institute of Philosophy	80,780
	MSU	Masaryk Institute	1,820
	PSU-P	Institute of Psychology	7,280
	SOU	Institute of Sociology	sect. lib.
		Size of Database	109,880
<b>MAZANKA</b>			
	UTIA	Institute of Information Theory and Automation	30,030

Server	Abbreviation	Name of Institute	Size of Library	
<b>MAZANKA</b>	OU	Oriental Institute	195,630	
	UIVT	Institute of Computer Science	8,270	
	UMCH	Institute of Macromolecular Chemistry	28,300	
	UFCH	Institute of Physical Chemistry	20,020	
	URE	Institute of Radio Engineering and Electronics	14,880	
			Size of Database	297,130
<b>KRC</b>				
	FGU	Institute of Physiology	360,000	
	UEM	Institute of Experimental Medicine	9,920	
	UMG-K	Institute of Molecular Genetics	see FGU	
	FKU	Institute of Pharmacology	1,000	
	MBU	Institute of Microbiology	FGU	
	BU	Institute of Botany	82,400	
			Size of Database	453,320
<b>PROSEK</b>				
	UTAM	Institute of Theoretical and Applied Mechanics	9,870	
	HIU	Institute of History	221,830	
	UZFG	Institute of Animal Physiology and Genetics	6,630	

Server	Abbreviation	Name of Institute	Size of Library
<b>PROSEK</b>	UZFG	Size of Database	238,330
<b>MACHA</b>			
	UEF	Institute of Ethnology	27,800
	UKS	Institute for Classical Studies	37,500
		Size of Database	65,300
<b>KLAROV</b>			
	ARU	Institute of Archaeology	64,400
		Size of Database	64,400
<b>SUCHDOL</b>			
	UCHP	Institute of Chemical Process Fundamentals	19,160
	UEB	Institute of Experimental Botany	19,460
	GLU	Institute of Geology	5,690
		Size of Database	44,310
<b>SPORILOV</b>			
	GFU	Institute of Geophysics	27,700
	UFA	Institute of Atmospheric Physics	7,200
		Size of Database	34,900
<b>PELLEOVKA</b>			
	UACH	Institute of Inorganic Chemistry	8,700
	UACH	Size of Database	8,700

<b>Server</b>	<b>Abbreviation</b>	<b>Name of Institute</b>	<b>Size of Library</b>
<b>TROJA</b>			
	A	Archive	98,230
		Size of Database	98,230
<b>REZ</b>			
	UJF	Nuclear Physics Institute	98,230
		Size of Database	98,230
<b>ONDREJOV</b>			
	ASU	Astronomical Institute	88,440
		Size of Database	88,440
<b>CESKE BUDEJOVICE</b>			
	ENTU	Institute of Entomology	
	HBU	Institute of Hydrobiology	
	PAU	Institute of Parasitology	
	STHSBP	Technical and Administrative Service of the ASCR Biological Center	40,250
	UMBR	Institute of Plant Molecular Biology	
	UPB	Institute of Soil Biology	
		Size of Database	40,250
<b>TREBON</b>			
	MBU	Institute of Microbiology	20,000

<b>Server</b>	<b>Abbreviation</b>	<b>Name of Institute</b>	<b>Size of Library</b>
<b>TREBON</b>	HBU	Institute of Hydrobiology	5,160
		Size of Database	25,160
<b>BRNO</b>			
	UIACH	Institute of Analytical Chemistry	6,000
	ARU-B	Institute of Archaeology	39,000
	BFU	Institute of Biophysics	7,850
	UPT	Institute of Scientific Instruments	19,950
	UEK	Institute of Landscape Ecology	32,660
	PSU-B	Institute of Psychology	4,330
	UEF	Institute of Ethnology	13,000
	UJC-B	The Czech Language Institute	18,800
	UCL-B	Institute of Czech Literature	15,000
	UFM	Institute of Physics of Materials	13,400
		Size of Database	163,990
<b>PARDUBICE</b>			
	SLCHPL	Joint Laboratory of the Solid State Chemistry	2,000
		Size of Database	2,000
<b>OSTRAVA</b>			
	UGN	Institute of Geonics	8,740
		Size of Database	8,740



## **Chapter 9**

# **CASLIN Uniform Information Gateway**

Bohdana Stoklasová and Pavel Krbec

At present, the majority of Czech libraries are hybrid libraries that provide information from both traditional and electronic resources and, in addition to their own information resources, rely to an ever-increasing extent on external domestic as well as foreign resources. The heterogeneous and international nature of information resources offers new possibilities that would have been difficult to imagine just a few years ago, but it also poses a number of problems that libraries must resolve in order to provide the maximum possible efficiency for their clients. Libraries should offer their clients integration of their services in a single user-friendly environment without the need to repeatedly log in and out, the ability to present queries in a uniform manner, to receive outputs, i.e. both information on documents and the primary documents themselves (most importantly full texts, but also graphics, sound, etc.) in a uniform format, and, on the basis of information thus obtained, to facilitate access to offers of further relevant information, and the possibility to work in the clients' own predefined environment with predefined preferred resources. In other words, most of what is so annoying to users today should take place 'in the background'. The CASLIN Uniform Information Gateway described in this paper provides the above functionalities. Since the CASLIN Uniform Gateway serves also as a virtual union catalog, these functionalities are described in the paper together with a possible model of cooperation of both real and virtual union catalogs under the umbrella of CASLIN.

## 1 The Beginning of the Project: Motivations

Information sources are not the only thing that has changed. The clients of libraries are also changing in their perceptions and in the demands they place upon libraries. Having access only to an online library catalog with a user-unfriendly interface will not placate them. They want much more. Being users of the Internet, they have grown accustomed to having immediate access, and they quite naturally also expect libraries to provide easy access to information. If libraries and librarians are to successfully compete on the information market, they have to offer qualitatively new services. At present, clients of heterogeneous libraries must be able to deal with different user interfaces (more or less friendly) of different services, learn a number of different query formats, repeatedly log in and log out, handle outputs of vastly different character, and resolve problems of different output formats and of different coding of diacritics.

In order to achieve integration of services in a single user-friendly environment, libraries need good-quality software providing the functionalities noted earlier, good-quality tools to manage information about both internal and external sources, tools for the management of information about users, and a sufficiently large staff of skilled librarians to effectively utilize all those tools.

An obstacle to full accessibility to the traditional holdings of Czech libraries and to easy navigation by users seeking information on a specific subject in libraries with particularly good resources in relevant areas is the poor quality of descriptions of the content of most Czech libraries' collections, and the absence of data for their overall viewing. What is missing is a comprehensive and easy-to-understand map of Czech libraries based on a common methodology. Coordination in the development and utilization of collections is not one of the strong points of the Czech library sector. Consequently, scarce financial resources for the purchase of documents are not used optimally. At the same time, Czech libraries have presented no convincing arguments that would enhance their chances for more money for collection development.

## 2 Project Objectives

The project Uniform Information Gateway for Hybrid Libraries<sup>1</sup> aims at improving the situation in all the above areas.

The objective of the project was to set up a uniform information gateway (UIG) that would allow users uniform and easy access to both traditional library holdings and local and remote electronic resources. The result of the project would be a gateway for the National Library as well as for Charles University, whose students and faculty it traditionally serves.

The project follows two main trajectories:

1. Implementation of foreign technical tools and standards for the UIG in the Czech Republic. Based on an analysis of the most appropriate tools for the attainment of the project objective, two products distributed by the Ex Libris company<sup>2</sup> were selected, namely MetaLib<sup>3</sup> and SFX.<sup>4</sup> The most important international standards used include OpenURL, Z39.50, UNICODE and MARC21; and
2. Determination of the prerequisites for optimum operation of the UIG, uniform subject cataloging, and a uniform description and analysis of Czech library collections based on the conspectus method and cooperation in their development and utilization.<sup>5</sup>

The UIG has been developing rapidly, and a number of changes have been implemented since the end of 2001. The present situation is described below.

---

<sup>1</sup> See <http://jib-info.cuni.cz/dokumenty/branaprojekt.html>.

<sup>2</sup> See <http://www.exlibris.co.il/intro1.html>.

<sup>3</sup> See <http://www.exlibris.co.il/metalib/overview.html>.

<sup>4</sup> See <http://www.sfxit.com/>.

<sup>5</sup> For results of the project in 2001, see <http://jib-info.cuni.cz/dokumenty/zprava2001/JIBTEXT.htm>.

### 3 The Present Situation

The objective of the original project was to establish the UIG for the National Library and Charles University, with the understanding that other libraries would be invited to participate in the project after it has been implemented and tested in those two institutions in the pilot stage. A number of libraries participating in a similarly focused project for the Science, Technology and Medicine (STM) division had already joined the project in the first year of its implementation (2001).<sup>6</sup> That made our joint effort a de facto national information gateway project from the very beginning. Such a project requires better hardware and software, and also more human resources, than a pilot project. This was the main reason for submitting the Library Public Information Services Program project.<sup>7</sup> In the year 2002, the project transcended national boundaries when Slovak libraries expressed an interest in participating in it.

### 4 Why the CASLIN Uniform Information Gateway?

The smooth start of the UIG project was made possible by the results achieved in previous years within CASLIN project activities<sup>8</sup> (especially in standardization and in setting up a uniform basis for a library network), the UIG being one of its logical outcomes, which is also the message on the UIG opening screen.

At present, the following Czech catalogs and databases are available to all MetaLib users through the UIG as freely accessible resources (sites that may be searched):

---

<sup>6</sup> See [http://jib-info.cuni.cz/o\\_nas/stm/jib-stm.html](http://jib-info.cuni.cz/o_nas/stm/jib-stm.html).

<sup>7</sup> See [http://www.nkp.cz/o\\_knihovnach/English/LPISindex.htm](http://www.nkp.cz/o_knihovnach/English/LPISindex.htm),  
<http://jib-info.cuni.cz/dokumenty/visk8projekt/visk8projekt.htm>.

<sup>8</sup> See <http://www.caslin.cz:7777/caslin/historie/document.html>.

**Table 1. Available Catalogs and Databases in MetaLib**

ANAL-články VK Olomouc (VKOL)	ANAL-článková bibl. (NK ČR)
AUT-báze autorit (NK ČR)	CASLIN-soub.katalog ČR
KKL-knihovnická lit. (NK ČR)	Katalog (MSVK Ostrava)
Katalog dokumentů (SVK Plzeň)	Katalog knih STK (STK)
Katalog-knihy (KVK Liberec)	MZK-katalog MZK Brno (MZK)
NFA-katalog dokumentů	NKC-katalog NK (NK ČR)
OPAC (Uk Upa)	SLK-katalog Slov.knih. (NK ČR)
SVK01-katalog VK Olomouc (VKOL)	Souborný kat.Univ.Karlovy (UK)

UIG users can also use a number of catalogs and databases of libraries abroad. The most frequently used US libraries are the Library of Congress Online Catalog, WorldCat (OCLC) and the University of California Digital Library. Records may be viewed separately after the appropriate source (e.g. WorldCat) has been selected. The Czech records come to WorldCat from the National Library of the Czech Republic, which, based on an agreement with OCLC, has been sending Czech National Bibliography records to the WorldCat catalog for a number of years, where they can be used by foreign libraries and their users. The application of AARC2R and LCSH in the Czech Republic makes mutual cooperation easier. At present, conversion from UNIMARC to MARC21 is necessary; in the future, Czech libraries plan to implement MARC21, which will make the situation even easier.

In recent years, records sent by the National Library to the WorldCat database have been equipped with subject headings (LCSH) in English, which is highly appreciated by foreign libraries:

Mi haj lo Rostohar (1878-1966) v tradi ci cel ostní a experi mentál ní  
psychol ogi e =

Mi haj lo Rostohar (1878-1966) frui tful tradi ti on of i ntegrate and  
experi mental psychol ogy / [uspořádal Josef Švancara] . -- Vyd.  
1. -- Brno : Masarykova uni verzi ta, 1999. -- 115 s. : il.,  
portréty ; 24 cm. -- -- (Sborník prací Filozofi cké fakul ty  
brněnské uni verzi ty ; P 3 (1999)

UDC: \* 159.9-051 \* 159.9.07 \* 159.9.019.2 \* 378.4 \* (066)

SH - Czech

Rostohar Mi haj lo, 1878-1966  
psychol ogové -- Sl ovi nsko  
psychol ogové -- Česko -- stol. 20.  
experi mentál ní psychol ogi e  
tvarová psychol ogi e  
uni verzi ty -- Česko

SH Engl i sh

Rostohar Mi haj lo, 1878-1966  
Psychol ogi sts -- Sl oveni a  
Psychol ogi sts -- Czech Republ ic  
Psychol ogy, Experi mental  
Gestal t psychol ogy  
Uni versi ti es and col leges -- Czech Republ ic

Figure 1. Illustrative Headings

Thanks to the cooperation with OCLC and the credit received for Czech records, Czech libraries can afford to use WorldCat records, which replaces original cataloging of their foreign acquisitions. Our cooperation with the WorldCat international union catalog is well known worldwide. Our foreign colleagues are often surprised and confused by the fact that our shared cataloging at national level is much less successful.

The sources that have been selected for the UIG out of the several hundred ones available are those where a high level of use by Czech libraries is expected. Other sources will be added according to requests from the UIG clients.

After a registration and necessary verification, other (paid) sources, including full texts, are available to the users. Table 2 shows what is currently offered in the subject category Economic Sciences and Business.

**Table 2. Economic Sciences and Business**

ABI/Inform (ProQuest)	ANAL-články VK Olomouc(VKOL)
ANAL-článková bibl.(NK ČR)	AUJ-jmenné autority (NK ČR)
Academic Search Premier(Ebsco)	ArticleFirst (OCLC)
Academic Source Premier(Ebsco)	CASLIN-soub.katalog ČR
DANBIB	Ecollections (OCLC)
IDS Basel/Bern	IDS Luzern
IDS NEBIS	IDS St Gallen
IDS Zurich Universitet	Katalog (MSVK Ostrava)
Katalog-knihy (KVK Liberec)	Library of Congress Online Cat
MZK-katalog MZK Brno (MZK)	MasterFILE Premier (Ebsco)
MZK-katalog MZK Brno (MZK)	NetFirst (OCLC)

ANL FULL-plné texty (NK ČR)	Acad. Research Lib. (ProQuest)
Account. Tax (ProQuest)	Asian Business (ProQuest)
Banking Inf. Source (ProQuest)	Business Wire News (Ebsco)
Báze české literatury (STK)	EconLit (SP)
European Business (ProQuest)	GEK-generální katalog (VKOL)
General Sci Plus (ProQuest)	ISSN (STK)
Journal Citation Reports (ISI)	KZP-zahraniční period.(NK ČR)
Katalog germanik (KVK Liberec)	Katalog-seriály (KVK Liberec)
Katalog-články (KVK Liberec)	Know@urope (ProQuest)

The extended services feature (SFX) permits UIG users to navigate from the source to other related targets.

## 5 Conspectus

Another important aspect of the UIG project is the development of prerequisites for its optimum operation: uniform subject-cataloging, use of the conspectus for a uniform description of the collections of Czech libraries, and cooperation in their development and utilization. We shall discuss this aspect only briefly.

The prerequisite for uniform subject-processing is the establishment of a national standard, i.e. the subject authority file. The subject authority file is gradually being built and published at the NL. It is based on the Library of Congress Subject Headings international standard (LCSH). Authority records include a notation symbol for the systematic selection language (Universal Decimal Classification) connected with the authority heading. This creates a connection between the subject and the systematic selection language for greater user satisfaction in searches.

Comprehensive accessibility of collections of Czech libraries and easy navigation by users are hindered by the poor quality of collection descriptions on the websites of a majority of Czech libraries, and by the absence of suitable data on collections in our libraries. There is no comprehensive and easy-to-understand thematic map of Czech libraries based on a uniform methodology, and coordination in the development and utilization of collections is not among the strongest points of the Czech libraries either. The Czech libraries' use of the conspectus approach (developed in the USA) should contribute to improving the situation.<sup>10</sup> The conspectus approach has so far been applied in the Netherlands, and the results and the necessary documents are available on the National Library

---

<sup>9</sup> See the OCLC/WLN Collection Assessment and Analysis Service, <http://www.oclc.org/western/products/aca/conspect.htm>.

<sup>10</sup> See <http://www.wln.org/products/aca/conspect.htm>.

website.<sup>11</sup> The conspectus subject categories were applied in the UIG project where they constitute ‘thematic crossroads.’

## 6 MetaLib and SFX, or the Basic Software for the Project

The most primitive search tools are simple information portals. They are basically lists of heterogeneous sources (sometimes thematically structured or thematically oriented), and the user merely selects the source to be used. Users will proceed differently with different sources because their formats for both queries and answers are typically different. A typical example is the ‘Information Gate’ at Charles University (which, in fact, is a portal and not a gateway, at least according to our definitions stated below). For a specific group of users (those who have an IP address from the Charles University block), the portal offers specific (and always identical) information resources. Unauthorized clients cannot use the portal. The real situation is somewhat more complicated, but this characterization will suffice for our purposes.

Parallel browsers are products of the gateway type that are more sophisticated in certain respects. They can send a query to several targets, and then use a uniform format to present the answers. The simplest parallel browsers will carry out searches in databases of the same type and, at the same time, provide the necessary interface. Parallel searches can be carried out in all databases equipped with the same browser/interface.

We may simplify this a little by saying that when parallel browsers are used, the communicating systems have the same interface, the sources are homogeneous, and the peer-to-peer communication model is used. Although services are offered to everybody, they need not be the same for everybody, and some of the services may be reserved for specific users only.

It would, of course, be convenient to have a tool that would have the characteristics of both a portal and a parallel browser. Such a tool would

---

<sup>11</sup> See <http://www.nkp.cz/konspekt>.

- offer access to heterogeneous sources as a portal, but always (from the user's point of view) in the identical way (as in the case of a browser);
- present results always in an identical uniform and easy-to-read manner; and
- distinguish between individual users and offer them user-tailored options.

We are going to refer to such tools as information gateways. It is obvious from the requirements imposed on information gateways that each information gateway must have instruments for the management and description (i.e. cataloging) of information resources, and also instruments for the management and description of the users and their rights. It must also have means for communication with the sources. It might therefore be tempting to build information gateways using the same elements that are used for the construction of library systems, which is what Ex Libris did by using the principles and technologies applied in its ALEPH library system to build its MetaLib information gateway.

Every information gateway (including also MetaLib), at least by our definition of information gateways, is an intelligent parallel browser in heterogeneous information resources. Search is the only service specified so far. But users know that browsers (OPACs) of library systems offer more services than just search (some services related to circulation). It would therefore be appropriate and satisfying to have information gateways that would also offer some additional services. However, it appears that it is advantageous to operate the 'system for the provision of extended services' separately from the search system, to make sure that the two systems can cooperate with each other, and to make it possible for the 'extended services system' to operate independently of the search system, that is to say independently of the information gateway.

An autonomous system for the provision of extended services has been developed at Ghent University. It is called SFX, which stands for Special Effects. SFX was developed by Herbert Van de Sompel. Ex Libris bought the SFX system and has been developing it further since. In combination with MetaLib (but also independently), SFX is a tool that can significantly enhance the productivity of work with heterogeneous information resources in the Internet environment.

The way that MetaLib and SFX work is that MetaLib connects to the Universal Gateway, which in turn connects to ALEPH, Z39.50, HTTP and

other clients. These in turn address diverse information sources. It is obvious at first sight that conceptually, MetaLib is not very different from a parallel browser. It does, however, differ in its scope and in the universal character of resources and interfaces allowed by it. MetaLib can search not only catalogs, but also all usual information resources, and it is not limited to any predefined interfaces. It, of course, uses Z39.50 for communication, but it can also communicate using its own interface. The typical resources for searching include catalogs, full texts, databases and archives. The services offered by MetaLib are user-tailored, and the system must therefore have the means and data necessary for the authentication and authorization of its users. (By authentication we mean checking the user's identity by means of a login and a password, or, alternatively, assigning the 'anonymous' status. Authentication is used for opening of specific personalized profiles. Authorization is a process by which, on the basis of the login and the IP of the user's address, his/her status, home institution, etc., access rights to resources are allocated in accordance with internal definitions.) Authorization is resource-related, and the system must therefore also maintain and manage data on resources. All the above data are put together in the so-called KnowledgeBase.

The KnowledgeBase also includes a description of processes such as resource handling. MetaLib will typically rephrase queries into a format that is appropriate for the resource selected, will send the queries and receive answers (results), transform them into its own format and output them. It will offer deduplication and if requested, will perform it. If it is operated in conjunction with SFX, it will also offer extended services.

MetaLib basically gets data, analyzes data and presents them, or jumps to provide extended services. As a result, the concept of MetaLib is simple. However, the SFX system for extended services is complicated. In order to gain at least some insight into the way it operates, we will need to dynamically differentiate the entities of the information world according to what role we assign to them at any particular moment. The entity through which we have just made a search, i.e. the entity we are in, will be called the source. After the search is completed, the source may offer extended (additional) services to us; the simplest service is constructing a hyperlink, i.e. taking a step aside. The location or the entity where the service is being provided is called the target. SFX can then be characterized as a system

providing and coordinating cooperation between sources and targets. SFX may be visualized as an 'observation tower' from which one can reach a variety of resources such as OPACs, ILL, full text resources, Web resources, citations, etc. Let us assume that MetaLib (which is one of the possible SFX sources) has found a record and used an SFX icon to offer extended, and for the moment unspecified, services. By activating the service (by clicking the SFX icon), the user will generate a source (i.e. MetaLib) request and will send it to the SFX system (to the SFX server). The request is in the so-called OpenURL format, and it contains record metadata, user identification, and source identification only. Hence, OpenURL does not contain any data about targets. The SFX server will process the OpenURL (it is, just like MetaLib, equipped with a knowledge base), and will offer concrete extended services to the user according to the record (which was the reason for sending the metadata) and the user (which is why the identification data were sent). The user then may decide to activate one of the services.

What needs to be done before a specific system can actively use MetaLib, and perhaps also function as an SFX source and/or target? As a first step, Czech (and Slovak) libraries would need to contact the National Library, and then, for technical details, the Computer Science Center of Charles University. The connection with MetaLib can be accomplished more quickly and easily if the system to be connected has the Z39.50 interface, but other possibilities also exist. While only systems that are able to generate the OpenURL can be SFX sources, practically any system can function as an SFX target.

## 7 Outlook

By the end of 2002 full Czech and English versions of the UIG will exist. It is expected that most of the major Czech libraries (regional libraries, central specialized libraries plus union catalogs of universities), some Slovak libraries (the virtual Czech and Slovak union catalog CASLIN), most of the STM libraries, and a number of other foreign libraries and resources will be actively participating in the project (that is to say, will make their resources available to it). Their number in the original

knowledge base is constantly growing, and their selection for the UIG will depend on users' interest in individual resources). Besides foreign full-text resources, domestic full-text resources will also be included (negotiations about required standards are underway).

The copy cataloging functionality will be made operational not only for ALEPH, but also for other library systems used by Czech libraries. The development of the subject authority file of the National Library will continue, and experience from this area will continue to be provided to other libraries.

Application of conspectus in the National Library will continue, and experience from this area will continue to be provided to other libraries.

A series of one-day seminars on UIG has started in the National Library training center. Other workshops will be organized as a part of the project STM Portal and at regional level. Presentations of the UIG will be made at national events (Automation of Libraries, Contemporary Libraries, RUFIS), abroad (Conference on Union Catalogs in Tallinn), and in print media (media from the above events, *Národní knihovna*, *Knížnica*, and other journals abroad have shown interest); at the end of the year, a short monograph in Czech and in English will be published at the conclusion of the R&D pilot project.

UIG financing will be provided through the VISK governmental grant program, and it will become a broadly used and indispensable tool for both users of libraries and librarians.

At some as yet undetermined later time, more resources and institutions will be included.

The copy cataloging functionality will be extended to include the format selection option for records copied (UNIMARC, MARC21), which will facilitate the use of international resources and the transfer of Czech (Slovak) libraries to the MARC21 format.

The existing functionality will be continuously enhanced and extended.

The conspectus concept will be implemented in a number of Czech libraries (in addition to the National Library).

Czech libraries will start cooperating in the development of their collections and in the building of thematic gateways.

Training sessions on the UIG will continue at different levels. Presentations of the UIG will be made at important events at home and abroad (IFLA conference in 2003).

Financing of UIG operations will be provided via VISK, or via conjunction with other national programs of the Ministry of Education and other ministries. The UIG will become an even more utilized and indispensable tool for both library users and librarians, and it will be also introduced to some non-library environments.

## 8 The CASLIN Uniform Information Gateway and the CASLIN Union Catalog of the Czech Republic

The parallel existence of UIG, which, as discussed above, is among other things also a virtual union catalog, and of CASLIN, a real union catalog, frequently raises questions such as: which is better, the real CASLIN union catalog or the virtual UIG union catalog? Do we need UIG now that we have the CASLIN real union catalog (and vice versa)? Will we need the CASLIN real union catalog once UIG has been put in full operation?

To begin with, it should be made clear that this is not a question of competition or a fight for a 'place in the sun' between CASLIN UC and the CASLIN UIG. On the contrary, it is necessary to make an all-out effort to ensure that the two systems operate smoothly and complement each other. In an ideal situation, UIG would be linked to a well-functioning real national union catalog and other union catalogs, including foreign and international ones. In certain respects, the UIG offers greater opportunities than a real union catalog. However, in other respects it offers fewer. With respect to resources, it offers more than CASLIN UC, thanks to its direct integration of foreign resources and extended services. But because of its broad sweep, it cannot provide for a comprehensive integration of all small Czech libraries. This can be done much better in CASLIN UC. With respect to functionality, the UIG's advantage is that it can search for and localize documents in libraries both in the Czech Republic and abroad, down to the level of current status of library items. Catalogers using the records downloading function will certainly appreciate the ease of selecting (or even permanently preselecting) the libraries, including foreign libraries,

from which the institution in question wants the records to be downloaded. For their various departments, institutions may even define different preselected menus (i.e. lists of institutions) for records downloading. The choice of formats (UNIMARC or MARC21) will make it easier for Czech libraries not only to use foreign resources, but also to switch from UNIMARC to MARC21. The UIG does not, however, nor will ever, provide for online cataloging into a common database. The absence of a common physical database of bibliographic records is an essential characteristic of virtual union catalogs.

When the real union catalog is placed under ALEPH, Czech libraries will have both the real and the virtual catalogs available in smoothly cooperating software environments. ALEPH and MetaLib have both been developed by Ex Libris, and that common cradle is apparent. It will be up to us to put that advantage to use. Clearly, good software support in itself is no guarantee of successful implementation and of frequent use of a union catalog. Issues that need to be carefully considered with regard to the development and use of union catalogs include strategic and conceptual ones. In many libraries, work processes will also need to be carefully re-evaluated, and in many cases substantially changed, to provide for a purposeful integration of the development and use of union catalogs into these processes. For several years, it seemed that the main problem lay in the shortage of technologies for cooperation within the union catalog. Of course, it is much easier to avoid cooperation when the technical tools available and needed for such cooperation are not on a par with those that the potential cooperating libraries are used to. However, with improvements in these tools, it becomes more and more obvious that the management of Czech libraries will have to surmount a much more difficult obstacle: namely, the natural human resistance to change.

## 9 Conclusion

By the end of 2002, good-quality virtual and physical union catalogs will both be available to Czech libraries. The two catalogs will complement each other. It has seemed that the main obstacle to shared cataloging is the imperfection of technical tools. But technical tools are getting more and

more sophisticated, and a significantly more serious challenge is posed by the lack of willingness to cooperate and substantially transform work processes in Czech libraries. May one hope that the situation will get better in the foreseeable future? The solution may come from an integration of electronic information resources with library services. Although UIG seems to be an extremely powerful technical tool, the full integration of electronic information resources to library services will also require human resources.

To generate resources for such integration, Czech libraries will have to reorganize and streamline their work processes, just as libraries in other countries did, and they will have to muster the will to cooperate and share responsibilities. The fact that user and economic pressures are markedly more subdued in the Czech Republic than in others is no advantage: because we are not forced to cooperate and share responsibilities, a lot will depend on our ability to self-start.

Integration of electronic information resources is a challenge that libraries abroad have had to deal with in recent years, and Czech libraries will have to do the same. We have so far been able to turn a blind eye to the issue of full-scale integration of electronic sources, because the pressure from library users has not been as great as in some other countries. We keep on saying that we will go ahead with integration when the time is right and when we have sufficient resources. We ignore the fact that the time was right long ago, and that processing, storing and providing access to electronic sources (including remote ones) has for a long time not been a luxury, but an absolutely standard library activity. This is underscored by the agenda of library conferences, seminars and workshops, or the websites of some libraries in other countries.

All of us like to process and provide access to classic documents. We have been doing that for long, we are accustomed to it, and we know how to do it. The regular arrival of a certain number of books for complete, original cataloging gives everybody a feeling of pleasant certainty. This feeling is even stronger when catalogers have a certain amount of backlog. We do not care that in Czech libraries one and the same book is processed many times. Their procedures are set, and shared cataloging means an unwelcome interference and a loss of that certainty and a loss of splendid isolation.

We do not care that Czech electronic resources are not processed at all and are irrevocably lost, and that neither contemporary nor future users will be able to access them. Expensive foreign resources lie idle in many Czech libraries, insufficiently advertised and utilized. Sometimes libraries themselves cannot use them, and cannot advise other users either. They say they do not have the time for these luxuries.

Integration of electronic resources is no luxury. Allocating vast resources to multiple activities related to classic documents, and especially to their repeated cataloging, is a luxury. It is unhealthy and untenable when, in one area, everybody does what everybody else does and could do better cooperatively, while another, equally important, area remains a no man's land. This observation gives us the hope that shared cataloging will be introduced on a nationwide basis.

### **Selected Bibliography**

Stoklasová, Bohdana and Pavel Krbec. Jednotná informační brána: projekt Národní knihovny ČR a Univerzity Karlovy. Paper presented at the INFORUM Conference, Prague: May 2002.

<http://www.inforum.cz/inforum2002/english/prednaska5.htm>,

<http://jib-info.cuni.cz/dokumenty/inforum2002/Inforum.doc>,

<http://jib-info.cuni.cz/dokumenty/inforum2002/Inforum.ppt>.

Stoklasová, Bohdana. "Jednotná informační brána CASLIN." *Ikaros* [online], (2002): č. 1 [cit. 2002-02-01],

<http://ikaros.ff.cuni.cz/2002/c02/jib.htm>.

Stoklasová, Bohdana. "Konspekt jako nástroj pro popis a hodnocení fondů." *Ikaros* [online], (2000): č. 10 [cit. 2000-12-01],

<http://ikaros.ff.cuni.cz/2000/c10/konspekt.htm>.

See also <http://jib-info.cuni.cz>.



# Chapter 10

## The Slovak Union Catalog for Serials

Lýdia Sedláčková and Alojz Androvič

In practical terms the union catalog is far from  
obsolete [...]<sup>1</sup>

The free flow of information and knowledge is a basic prerequisite for the development of modern societies, and is exemplified by a united Europe and other advanced societies around the world. The coordinated creation of and access to library catalogs, relying on modern technology, make significant contributions to those societies' development. Long-term practice supports the belief that one of the most effective instruments for promoting the free flow of knowledge is the union catalog and the best method for creating it is cooperative cataloging, which is labor-saving and contributes to the quality and speed of cataloging.

A national union catalog is the fundamental information resource for documents in the libraries of a country. One of its important functions is its ability to locate information. In addition to its cataloging functions, it standardizes and stores information and provides opportunities of cooperation and coordination. The union catalog contains the holdings of the participating libraries as well as the national document production. It reflects the culture and cultural heritage of the country in question. Universal union catalogs incorporate a range of processes for registering,

---

<sup>1</sup> Clifford A. Lynch, "Building the Infrastructure of Resource Sharing: Union Catalogs, Distributed Search and Cross-Database Linkage," in *Souborné Katalogy: Organizace a Služby*, Prague: Národní knihovna, 2000: 21.

preserving and presenting information about national and international intellectual output.

New technologies have extended the scope of union catalogs by enabling the inclusion of international library materials and electronic transmission of the articles from periodicals.

The construction of a Slovak union catalog for periodicals (UCP) has been an old tradition at the University Library in Bratislava. The creation and improvement of union catalogs for periodicals and of other extended services continue to be important and hotly debated issues in many countries and regions.

## 1 The UCP in a Changing World

Once, the union catalog was kept on catalog cards; it<sup>2</sup> described the collection holdings of a number of libraries.

The present solutions applied in the Slovak Union Catalog of Periodicals (UCP) have not been invented ‘on the fly.’ The culture of union catalogs (UC) is deeply rooted among Slovak librarians, as it is in many other European countries. This familiarity with the phenomenon was certainly important insofar as it offered a foundation from which to build a new system. It meant that we were aware of the complexity of the project, of the potential pitfalls and of all that is involved in bringing such a project to successful completion. On the other hand, this past history could also become a burden as it limited our ability to imagine our own goals and visualize alternative solutions.

Slovak librarians realized the significance of union catalogs as early as the 1920s.

The activities of the first director of the University Library in Bratislava (ULB), Dr. h.c. Jan Emmeler, aided scientific and cultural developments in the former Czechoslovakia. His conviction that common catalogs are fundamental preconditions for scholarship led him to the idea of creating a

---

<sup>2</sup> Ole Husby, “Real and Virtual Union Catalogs,” in *Souborné Katalogy: Organizace a Služby*, Prague: Národní knihovna, 2000: 112.

union catalog. In 1923, during the international library congress in Paris, he presented the first report on the preparations for the Czechoslovak union catalog. In the same year, he designed a precise specification for the technology, which is still of value. It contained practically all the necessary ingredients for a modern union catalog, including the catalog type (alphabetical, bibliographic record card catalog); the role of the major libraries and the tasks they would have to carry out, the establishment of the union catalog's central database; the unification of the card format according to the international cataloging card format accepted by the International Bibliographic Institute in Brussels (12.5 x 7.5 cm); a library identification system and recording format for location and holdings information; a proposal to establish an expert Union Catalog Committee; a data flow scheme for the whole system (including the definition of the first record for the then International Bibliographic Institute in Brussels) defining the reference and bibliographic sources for the catalog; and cataloging instructions and the proposal of an agreement on common rules for lending (the first international recommendation for building union catalogs as a precondition for ILL was approved by the international library congress in 1935!).

The catalog concept was realized in Slovakia during the period 1923–1936. A central catalog of Slovak libraries was created, containing records of the rare collections in 13 predominantly historical libraries, founded and developed since the middle of the sixteenth century. The highly professional bibliographic work on the first union catalog became the best avenue to a retrospective national bibliography. During the 14 years of collective work, the union catalog grew to 50,000 records of historically valuable books and serial documents of various provenances.

In 1947, the reputation and the results of the project attracted the attention of Dr. Besterman, a representative of UNESCO. He visited the University library and presented a proposal for cooperation on a central catalog of UNESCO countries. However, the political situation was not favorable for such a project at that time.

However, the past efforts of building a union catalog in Slovakia were not ignored even after the changes in the political system in 1948. In 1949, official bodies selected the union catalog as one of the main tasks of the national library system. ULB was entrusted with a pivotal role, and in

consequence, the state expected reasonable growth in cultural life. However, there were many other more important political priorities at that time, such as the liquidation of monastic libraries, and for this reason union catalog production was postponed until 1957. Thereafter, the creation of the central union catalog began in three major libraries in parallel.

The production of union catalogs in the country and the obligations imposed on ULB were significantly impacted by Law No. 110/1965, which pertained to cataloging foreign literature. This legislative act required libraries involved in state-wide cooperation to create union catalogs and to serve as the basic information resources for inter-library loan services, for the cooperative provision of the acquired literature and for the accompanying financial evaluation of state resources used primarily for collection building. ULB had played a central role in this activity. All Czechoslovak libraries collecting foreign literature started to build cooperatively a Czechoslovak union catalog, divided into two parts:

1. A union catalog of foreign books (produced in parallel at National Library in Prague and at ULB)
2. A union catalog of foreign periodicals (ULB).

The common Czechoslovak state-wide foreign literature union catalog was constructed by 1993 (the year in which Czechoslovakia split into two countries). The union catalog of books still has foreign (including Czech) books in its scope and has a classic card catalog. It is still being added to, and continues to be used (it currently contains 3.6 million records and has 200 contributing Slovak libraries).

The foreign periodicals union catalog has passed through several stages of development. Towards the end of the 1960s and the beginning of the 1970s, ULB and the National Information Centre in Prague (NIC Prague) developed an automated periodicals union catalog (ASKKP), at first as an offline system. ULB, as the producer of this system, was responsible for gathering data, central administration and data conversion (to magnetic tape), as well as for the typographic processing of the printed version of the catalog. ULB published and distributed the catalog for the whole of Czechoslovakia.

The UCP database was maintained and updated in Micro-CDS/ISIS. It was not the only way to save the data, but luckily it also turned out to be a

very flexible and reliable environment for continuing the work. Even today, we still have the free Micro-CDS/ISIS package, which is a useful tool for data management. NIC in Prague developed a complex technology required for processing input data, formal and logical data control, and list editing. Since 1982, it has provided online dial-up access to the union catalog database (ASKKP), together with access to the address database of contributing institutions. Access to these model Czechoslovak bibliographic databases was realized through an international network connecting Moscow, Prague and Vienna. In a 1991 survey and study of the library and information system in Czechoslovakia, the British "Know How Fund" favorably evaluated this online periodicals union catalog database. The microfiche edition of the catalog was produced in parallel with the computerized database.

During the many years of their existence (until 1996), the online periodicals catalog database and the directory database have had several different structures, used several retrieval systems (GOLEM, STAIRS/CMS, Micro-CDS/ISIS) and several servers (IBM mainframe, mini-computer, PCs). In principle, the only invariant in this varied development was the serial publication itself and the bibliographic data about it. The database contained over twenty thousand titles of foreign periodicals in more than 1,400 Slovak and Czech libraries from 1976. Until 1996, the database was updated only once a year.

In the period 1991 to 1995, the ASKKP database and the address database were also available on the Slovak academic network SANET, connected to the Internet and maintained in the STAIRS system at the Institute for Applied Information Science in Bratislava. Libraries and other users were able to copy selective outputs from the union catalog database to floppy disks for use in the local computers that were also using the UNESCO system Micro-CDS/ISIS.

## 2 CASLIN: A Milestone of Czech and Slovak Librarianship

The contrast between the trend toward globalization and the efforts of individual nations to maintain their national and cultural identity has increased the importance of integration, cooperation, standardization

and harmonization. In Slovakia, this contrast has raised the importance of the Czech and Slovak Library Information Network (CASLIN), founded in 1991.

Four main Czech and Slovak libraries, the Czech National Library, the Slovak National Library, the Moravian State Library in Brno and the University Library in Bratislava, agreed to create a solid foundation for a nationwide library network. Slovak participation in this network increased when the East Slovak library consortium KOLIN joined CASLIN.

The network was designed as an integrated cooperative system based on shared cataloging and the utilization of central processing of the national production of library materials. One of the main tasks was the gradual construction of a union catalog of all participating libraries. Czechoslovak union catalogs of foreign literature became the basis for this union catalog.

The standards adopted in CASLIN include

- the exchange format UNIMARC;
- international bibliographic recommendations for bibliographic description, ISBD;
- Anglo-American Cataloging Rules, AACR2;
- Universal Decimal Classification as a classification scheme.

The technology includes

- the integrated library system ALEPH;
- an Internet network environment.

The adoption of international standards and rules allows for common bibliographic descriptions and provides for a national and international record exchange system. At present, Slovak libraries use ISBD standards, AACR2, UNIMARC format, and the ISO 2709 standard for bibliographic records exchange in electronic form. The application of the above standards is obligatory for the union catalog.

In the initial project years, all participants concentrated on building their OPACs, their locally produced and maintained catalogs of monographs. The idea of the union catalog was more theoretical than practical, particularly because of weaknesses in the telecommunications infrastructure. It was also the case that the particular version of the software acquired did not lend itself easily to implementing shared cataloging. But the ULB created and

extended the UCP in accordance with all standards, rules and system procedures implemented in CASLIN. A copy of the UCP was provided to the National Library of the Czech Republic in Prague in 1991.

The first challenge was how to adapt the periodicals union catalog to the CASLIN standards, namely UNIMARC, ISBD (S) and AACR2. After the data structure comparison, we had provided several conversions and data modifications, including the specific holdings structures. Minimal and standard formats and data structures for serials and indexing were adopted. A significant change was introduced into the structure of holdings information. Finally, the Czechoslovak database was removed, thereby creating a purely Slovak pool of information.

The conversion of the UCP database into UNIMARC format and the customisation of the ALEPH system had been achieved in cooperation with the Czech National Library and since the end of 1996, the catalog had been accompanied by a complementary ADR database: (the directory of participating institutions) available via Internet.

The UCP and the participating libraries (ADR) relied on 2 parallel databases:

- a working cataloging database (micro-CDS/ISIS)
- a public user database, used for data retrieval (ALEPH 3.25).

In the cataloging database, the following functions were performed:

- cataloging of new serial titles;
- data update and correction;
- duplicate control;
- data export for local catalogs of participating institutions;
- data export for the public user database;
- export data for the German document delivery system JASON (Journal Articles );
- printout of the address directory.

The organizational and technical conditions have determined the composition of services offered to professional staff and end-users. Currently the following services are offered:

- Information retrieval in the UCP database and ADR database (the directory of participating institutions) in the WWW environment;
- Parallel searches on an experimental basis in both the Slovak UCP database and the German ILL database system JASON, which is connected to the bibliographical article database JADE. This system relies on electronic document ordering from the North Rhine-Westphalian libraries and more than 20 Slovak libraries;
- Information requests can be made by telephone or in writing (including fax and E-mail) directly to the Union Cataloging Department of the ULB, and this service is available to the end-users; and
- Every 2-3 years a printed version of the database—ADR (directory of participating institutions) is published. In 1998, the union catalog of periodicals was extended by the inclusion of the records of Slovak periodicals, and of the complementary database of the participating libraries directory, which has been available on the WWW since 1996.<sup>3</sup>

CASLIN has had a positive influence on library automation, the standardization of data processing, cooperation, library management, and practically all daily library activities. In the last decade of the past century, significant changes have occurred in library automation. Step by step, the Slovak libraries learned to organize their work according to international rules and standards.

In the Slovak Republic, responsibility for the creation of union catalogs was divided between the University Library in Bratislava, which is responsible for the union catalog of periodicals, and the Slovak National Library in Martin, which is the administrator of the union catalog of monographs. This structure was approved by the CASLIN project directors in 1995, and was adopted by the then new Slovak library law No.183/2000.

---

<sup>3</sup> See <http://www.ulib.sk>.

### 3 New Directions

The first UCP installation on the basis of the ALEPH 3.25 integrated library system (1996) was implemented as a static bibliographic catalog, regularly updated in batch mode with no shared cataloging capability. For two years, this catalog was accessible online via remote terminals and has been available on the Internet for the past six years. The update and control was provided in the background, using the Micro-CDS/ISIS package and the workflow settled into a satisfactory routine.

As mentioned earlier, the ALEPH implementation originated in the CASLIN project more than six years ago. Since that time, the quality and functionality of the system have changed continually. The only results achieved thus far have been the centrally maintained online databases, regularly updating the UCP user database.

Over the ensuing years, the need for an organizational and technological rethinking of our setup became increasingly more obvious, and so a new concept was discussed and drafted and, while the implementation platform was not in doubt, our rethinking focused primarily on the UCP model architecture and the nature of the data processing workflow. With the natural improvement in our knowledge, there was a dramatic improvement in the quality of bibliographic descriptions, and with that a corresponding improvement in overall UCP consistency. The situation was ripe for a dramatic leap forward.

In 2002, the Open Society Foundation in Bratislava invited a proposal for modernizing the UCP. The proposal was prepared with reference to CASLIN and in close cooperation with the Slovak National Library in Martin. The main goal was to implement a cooperative cataloging system, and the execution was scheduled for 2002. UCP was inspired by the best foreign and domestic practices, based on recent standards and using the latest available ALEPH 500 system environment. It is noteworthy that ULB runs the local library system ALEPH 500 V.11 for cataloging and circulation and the new UCP developments have to be provided in parallel. The project proposal was accepted, and the project funded for one year by OSF Bratislava.

The strategic goal of the project is to turn the national union catalog of periodicals into a rich and accurate source of information about the

availability of periodical documents in Slovak libraries, to create an important resource and tool for cooperative cataloging, to provide for the retrospective conversion of local periodical catalogs, and to create a basis for cooperation in the area of periodicals acquisition in Slovak libraries.

The project had a number of detailed objectives, and was expected to yield numerous ancillary benefits. First of all, it had to establish a cooperative system for the union catalog of periodicals as an integral part of the library system in Slovakia. This system would then support the active participation of libraries in the union catalog. It was likely to result in a reduction of costs due to lessened reliance on original cataloging. The broad participation of libraries would then enhance the skills of librarians through the library system, and would also accelerate the processing of periodicals. The project would also create conditions hospitable to retrospective conversion, and generally improve the quality and accuracy of the database through increased reliance on international standards. Searches in the catalog would become more effective, and the project might well point the way toward other enhanced services. ILL services were likely to become more effective, the overall costs of acquiring materials would be reduced, and joint collection management would be enhanced.

It was expected that the operations of the union catalog would lead to a gradually expanding number of participating libraries, and their overall operations would be much improved. It was deemed sensible to build on the experience and procedures of CASLIN, and hence it was natural that the union catalog should rely on the ALEPH software. In fact, the CASLIN and KOLIN libraries are expected to play a key role in the union catalog and have much to say about the planned features of the system. The implementation of the union catalog uses the most up-to-date version of ALEPH 500 system, namely Version 14.2.

In March 2002, the vendor of ALEPH, Ex Libris, provided a test period for the installed Version 14.2, Patch 4. This version was installed on a separate SUN 450 server. The previous efforts had concentrated on the customization of internal data formats (UNIMARC) and staff and user interfaces. The translations of the system messages and templates, the parameter settings for system tables and various minor system adjustments required extensive calibration before the first UCP record could be viewed on the screen. The control parameters of the recent version, very different

than the earlier ones, had to be set ab ovo. Conceptually speaking, the UCP was designed as a set of interconnected catalogs (ALEPH libraries) with standardized data structures for bibliographic data, holdings information and authority records for both the UCP (UCP BIB) with library addresses (UCP ADR) and ISSN (ISSN BIB), and for publisher addresses (PUB ADR). The bibliographical structures of UCP and ISSN are identical. Special Micro-CDS/ISIS export print formats were designed for formatting both the UCP data and ISSN data for import to ALEPH. The conversions were validated by repeated iterated export and import between the two systems. The ALEPH ISO 2709 export was successfully imported using a special Micro-CDS/ISIS field selection table (FST) after the data had been validated using the OSIRIS controls.

The data from the library addresses database (ADR) were exported in a similar way for the standalone ADR database, and also in HTML format for use in the UCP presentation links. Clicking the library codes in the holdings information of the retrieved serial record generates a frame with the library profile, containing direct links to the local catalog or the library WWW site.

The heterogeneity of the automated systems used in local catalogs complicates the situation. Automation systems in use include VTLS, Rapid Library, Libris, OLIB and CDS/ISIS. Only six libraries use ALEPH. The participating libraries will use an ALEPH 500 client for cataloging and downloading data. The data transfer to local non-ALEPH systems is to be solved at workstation level by sharing the locally generated (downloaded) record structure with the ALEPH client. Because of the different working regimes and maintenance cycles of local library systems, the UCP implementation and production environment will use a separate server and separate ALEPH system.

#### 4 The Model of the Union Catalog of Periodicals: UCP

In environments where a fixed-scope union catalog needs to be presented to a large patron community as a basic, high-quality, highly available resource, it seems clear that with current technology, centralized union catalogs have major advantages both in function and in performance.<sup>4</sup>

UCP developed traditionally as a centralized union catalog model. The idea of this type of physical union catalog was subsequently adopted by CASLIN.

New IT technologies facilitate the construction of virtual catalogs. There are many successful virtual union catalogs. However, numerous comparisons and evaluations suggest that “neither of these approaches is panacea, however— both have certain pros and cons, which helps to make the decision which to adopt dependent on circumstances.”<sup>5</sup>

The characteristic of the UCP could be “concrete, downloading to local systems.”<sup>6</sup> Common problems of union catalogs, such as data heterogeneity, structural homogeneity and semantic heterogeneity, could be handled in this model quite successfully. Another advantage of the centralized model is that it is also appropriate for the special characteristics of periodical documents, such as temporary changes in periodical identification elements, the variety and variability of periodical titles, the need for complementary basic data elements (for example key title, abbreviated title, history etc.), and the necessity for permanent control and update of records even after the serial has ceased to exist.

A centralized model ensures the existence of a precise information resource for ILL, the presence of a unified standard data presentation, the availability of holdings information for several libraries in one database, the consistent interpretation of requests and data, the usage of unified query

---

<sup>4</sup> Lynch 127.

<sup>5</sup> Mathew J. Dovey, “So You Want to Build a Union Catalog?” *Ariadne*, Issue 23. See <http://www.ariadne.ac.uk/issue23/dovey/intro.html>.

<sup>6</sup> Husby 115.

methods, and the attainment of more precise search results. The data processing centre acts as an expert authority, guarantees union catalog tasks and functions, ensures catalog integrity, reduces levels of record duplication, and also reduces the need for multiple control and supervision of the adopted rules. These advantages outweigh the disadvantages, such as the presence of outdated data in the catalog and the high costs of operating the central facility. A further reason for not phasing out a centralized catalog is the continuing poor level of Internet connectivity. In spite of the fact that the number of libraries connected to the Internet is increasing in Slovakia, there are still many without local periodicals catalogs and without connections to the Internet. Only a union catalog enables them to make their interesting serial collections available to the public at large.

## 5 Document Types

The structure of document types in the catalog is close to that in similar databases built in the German *Zeitschriften-Datenbank* (ZDB), and/or in the Austrian *Österreichische Zeitungen- und Zeitschriften-Datenbank*, (ÖZZDB), although due to their size, they are not really comparable. They offer coverage of the entire range of periodicals. They contain data on periodicals, journals, magazines, newspapers, yearbooks, proceedings, printed or electronic resources, microform documents, etc., with the exception of monographic series. New forms of publications, such as electronic documents, require new definitions of document characteristics. Some of the new document types are recognized by AACR. Changes in document types are reflected in the cataloging rules and international conventions and are accepted by large international systems, such as AACR2, ISSN and ISBD. What remains is the adaptation of these international systems to the national circumstances, a task that is being currently addressed.

## 6 Cataloging

One of the main goals of the new project is the cooperative cataloging of serials. In the model that is being implemented, records are kept in the central catalog and imported into local catalogs. Cataloging practice and

indexing policy are determined centrally, but they still allow individual members of the UCP to use their own systems. This practice still necessitates agreement among the contributing libraries concerning the cataloging policies used. A common structure does not achieve anything if there is no agreement on the contents of the structure.

ULB, as the administrator of the catalog, had accepted certain standards and rules. This has not been easy, because the UCP had exact rules for creating data, different from ISBD (S) and AACR2 in some cases. Since minimal record levels were accepted in 1996, catalog records have been processed in UNIMARC and bibliographical records follow the AACR2 and ISBD (S) rules.

Implementation of the cataloging rules that are new for us, such as AACR2 and ISBD (S), presuppose adequate preparation by librarians for handling bibliographic periodical data. In CASLIN, attention was focused only on bibliographic records of monographs. The cataloging of periodicals is more complicated, and cooperative cataloging requires the agreement of all parties.

Another difficult task is retrospective cataloging, and it would be desirable to utilize records prepared by librarians in other countries or data from the ISSN database.

The cataloging environment has become global, and cataloging discussions at international level have intensified during the last decade. Large cataloging agencies and information communities are discussing or carrying out revisions of their rules (AACR2, ISSN, ISBD, German RAK, Italian RICA). The results of this process will be challenging. Different cataloging rules create barriers. Harmonization of the different cataloging codes is affected by another essential factor, the functional requirements of the bibliographic record (FRBR), which has contributed to a theoretical understanding of the cataloging activity among cataloging concerns around the world. FRBR does offer a conceptual framework that has the power to bring different cataloging codes closer together and thus promotes compatibility.<sup>8</sup> Harmonization of the cataloging codes would be another step

---

<sup>7</sup> Husby.

<sup>8</sup> Husby.

to the utopian goal that a bibliographic resource shall only be described once, which would eliminate conversion discrepancies and problems between different countries and systems. Last but not least, it could have an economic impact as well. This is also a global problem. The activities in the past decade of the major players, such as the Library of Congress, the British Library and others, have not contributed to the consolidation of the scene. A set of national MARC clones has been maintained for a long time. The UNIMARC initiative did not find the necessary support even from its originators. And finally, there emerged the constructive idea to adopt MARC21.

## 7 Holdings

Important holdings and local information is to be found in every centralized union catalog. Considering that UNIMARC does not have any format for holdings, field 910 was defined for this purpose. In ALEPH 500, it is already possible to use holdings records. Complete holdings records will be provided by the MARC21 format. The structure of holdings also has an important impact on ILL. Syntactical analyses of the holdings structure in UCP have confirmed their 'readability.' This means, however, that rules that govern their creation must be uniform and clear.

## 8 Cooperation

The basic principles applied in UCP are continuity and accessibility. The continuity principle has been satisfied by the connection of UCP with existing foreign periodicals catalogs. UCP is open to all Slovak as well as foreign libraries, and may cooperate with other similar systems. More than 350 Slovak institutions provide data and cooperate in building the foreign periodicals catalog; the cooperating libraries include all scientific and academic libraries, public libraries, medical libraries and the Slovak Academy of Science libraries, libraries of enterprises of various sizes and of research institutions, etc. After 1989, the number of libraries taking part in the UCP decreased due to the closing of many companies and research institutions.

The academic libraries are growing gradually. These institutions are the co-creators and main users of the catalog. Representatives of all-important libraries with significant participation in the catalog meet in the coordination committee and in working groups. For example, a commission was established in 1996 for minimizing the level of records for print periodicals in UNIMARC, and in 2001 for setting minimal records for electronic serials. Further steps were taken to increase the state of knowledge, the application of international norms and standards in catalog and bibliographic processing of periodicals, and the creation of catalogs for such documents. The apparent problem with using formats other than UNIMARC may be solved in the future by overlapping it with the MARC21 format.

Cooperation with German librarians in the JASON system at the University Library in Bielefeld is also successful. The association between German technology and data and the Slovak UCP has formed a unique information source and provides for the electronic ordering and supply of serial documents.

The UCP project is built on close cooperation between the central UCP administration and the National ISSN Agency. UCP uses this system and its data to a maximum extent. At this point, the ISSN identifier contains 75% of the serial records in UCP. The ISSN Slovak database is the basic source for processing bibliographic records and is available to all UCP participants in ALEPH. The ISSN system carefully controls the lifecycle of serials (predecessors, successors, variant forms), and, as an international system, is subject to multiple expert controls. In the period of 1998–2001, the increase in new Slovak records in the ISSN system (new titles, significantly changed titles) reached 10% of the whole, while less significant changes were recorded in another 10%.

## 9 Classification and Indexing

The CASLIN libraries have accepted Universal Decimal Classification as the basic classification system. UCP has been organized according to UDC from its inception. It was interesting to discover through a survey of periodical processing that only 16 out of 230 participants in the survey used

UDC. Another 22 libraries mentioned using the keywords, their own systematic classifications, subject headings or a thesaurus.<sup>9</sup>

These findings speak for themselves. It is obvious that the present hierarchical classification scheme, UDC, needs to be complemented with another classification, retrieval language or indexing scheme. One option is to use a controlled vocabulary. Such a decision is not easy to make. There is no classification and indexing system available for a universal union catalog, and yet a selected system must be appropriate for all fields of knowledge and must be acceptable to all libraries that are participating in the union catalog.

The most commonly used and widely accepted subject vocabulary for general application is the Library of Congress Subject Headings schema. It is universal controlled vocabulary. However, LCSH's complex syntax and rules for constructing headings restrict its application by requiring highly skilled personnel, and limit the effectiveness of automated authority control. Partial application of several classification and indexation systems at the same time would cause confusion, and decrease the accuracy of information and of navigation by users in the whole system.

## 10 Serial Cataloging Training

Librarians need to be trained continually for cooperative cataloging and international standards. However, there are no organizational arrangements for this in Slovakia. From 1999 until 2001, ULB organized workshops on UNIMARC and the creation of new records of print and electronic serials. Currently, an educational program is being prepared. Our purpose is to increase the pool of educated serials catalogers and to raise the quality of serials cataloging records that are contributed to a shared database. We have prepared basic serials cataloging workshops, starting with the definition of the serial, followed by concepts of original and copy cataloging. Classification and new trends in serials cataloging, such as

---

<sup>9</sup> MeSH, AGROVOC, EUROVOC; the thesaurus of The British Zoological Society; the international thesaurus used in Astronomy and Astrophysics Abstracts, etc.

cataloging of electronic serials, will also be covered. A session on MARC21 coding may also be included. In the training of librarians and paraprofessionals, we would like to utilize the experience of the effective and authoritative American serials union catalog program, CONSER, and its standardized materials, manuals, and training methods.

## 11 Conclusions

The union catalog of periodicals, UCP SR, could, step by step, become a quality cooperative system, having direct links to the national ISSN system for bibliographic registration and identification. Much work remains to be done, including building the authority files, defining solutions for holdings-data records, solving the problem of automatic data updating among UCP and local catalogs, preparing the librarians for effective cooperation, handling cooperative cataloging, preparing complementary programs, achieving smoother cooperation with non-ALEPH systems, switching UCP to deal with the full texts of serials, building the electronic ordering system, and solving the problem of the archival storage of UCP on other data media (CD ROM, microfiche).

The tasks are challenging, and put costly demands on librarians in information technology. Experience from some other countries suggests that it may be more advantageous to try to solve the technological problems outside the libraries. To be successful, the process will require continual inventiveness, endurance and cooperation.

## References

- Dovey, Mathew J. "So you want to build a union catalog?" *Ariadne*, Issue 23. <http://www.ariadne.ac.uk/issue23/dovey/intro.html>: March 2000.
- Husby, Ole. "Real and Virtual Union Catalogs." Prague: Národní knihovna, 2000: 112–120.
- Jonsson, Gunila. "The Basis for a Record in Major Cataloging Codes and the Relation to the FRBR." IFLA Council and General Conference, August 18–24, 2002.

Lynch, Clifford A. "Building the Infrastructure of Resource Sharing: Union Catalogs, Distributed Search, and Cross-Database Linkage." *Souborné katalogy: organizace a služby*. Prague: Národní knihovna, 2000: 121–130.

Pelle, Françoise. "ISSN: An Ongoing Identifier in a Changing World." *Serial Librarian* 41/3-4 (2002): 31–42.

Sedláčková, Lýdia. "Súborné katalógy v systéme sprístupňovania dokumentov." *INFOS '98*. Bratislava, SSK, 1998: 74–77.

Sedláčková, Lýdia and Androvič, Alojz. "Union Catalog of Libraries in Slovak Republic." *Souborné katalogy: organizace a služby*. Prague: Národní knihovna, 2000: 173–178.

Statement of Intent. In *CASLIN 1991–2001*. Bratislava: Univerzitná knižnica, 2002: 18–20.

**Table 1. Basic Statistical Data of the UCP**

Item	2002
Model	Centralized physical union catalog
Format	Online database including cooperative cataloging/copy cataloging
Automated system	ALEPH 500, v.14.2, patch 4 (working in the heterogeneous environment)
Bibliographic exch. format	UNIMARC
Standards, cataloging rules, indexing	AACR2, ISBD (S), UDC
Type of documents	Serials
Number of records in the catalog	38,000
Number of records added per year	1,500–2,500
Number of contributing libraries	350

Item	2002
Type of contributing libraries	All types
Subject coverage	All types
Data range	All date
Retrospective conversion	In preparatory stage
How many serial records contain an ISSN?	75%
Records may be searched by	All titles, ISSN, issuing body, country code, language code, code of the libraries, UDC, corporation, place of publication, keywords from all fields, system number
Availability of documents for loans	90%
ILL requests should be sent	To the libraries holding the item

## Part 3

### Polish Union Catalogs



# Chapter 11

## Are Our Union Catalogs Satisfying Users' Needs?

### Thoughts on the Evaluation of Union Catalog Projects

Błażej Feret

User satisfaction may or may not be directly related to the performance of the library on a specific occasion.

– K. Elliott

#### 1 Introduction

Planning the present paper, I thought that I would be able to survey user needs and satisfaction concerning union catalogs in different countries under the umbrella of The Andrew W. Mellon Foundation. The principal reason for thinking that such a survey would be desirable was the conflict between two separate union catalog groups in Poland with respect to the philosophy and rules and the extent to which the catalogs would be available to as many libraries as possible.<sup>1</sup> It was very tempting to determine whether the Polish union catalog NUKat, in its ultimately agreed shape, was meeting user needs and satisfying them, and to compare it with other union catalogs. However, this task proved to be very complicated. How can one measure user satisfaction? How could one find out what users need? The literature provides examples of user satisfaction surveys

---

<sup>1</sup> Richard E. Quandt, *The Changing Landscape in Eastern Europe: Personal Reflections on Philanthropy and Technology Transfer* (New York: Oxford University Press, 2002): 244–247.

concerning both general and particular library services.<sup>2</sup> In most surveys, authors ask users to indicate their satisfaction level on a closed, 3-5 point scale, e.g. “very satisfied”, “satisfied”, “not satisfied”.<sup>3</sup> This approach works very well for long-established library services and for users with a high level of awareness of the library services in question. For projects that are relatively new, such as union catalogs in post-Communist countries, the problem is not so simple. These projects started only a few years ago, and some are still in their initial phase. In many cases, the declared goals have not yet been achieved. The term ‘user satisfaction’ usually describes the effects of the project after it has been completed. But can we also talk about satisfying users’ needs or meeting users’ expectations at the time that the union catalog is designed? Should the reference time be ‘now’ for all projects? Or perhaps the project goals could be assessed in terms of user satisfaction as early as the time a union catalog is designed? Or perhaps it is simply too early in transitional countries for research on user satisfaction concerning union catalogs?

Another question is: who are ‘the users’ to be surveyed? Are they librarians or non-librarians? The two groups will certainly have different expectations concerning the project (in all phases), and would therefore express different levels of satisfaction. How can one find out whether there exists a need for some particular function in a union catalog if users have never used a union catalog before?

Due to all these uncertainties, I deferred carrying out a survey for the time being, and instead I decided to discuss some general problems related to the evaluation of project results. In this paper, I try to identify several methods for assessing the results of union catalog projects. I discuss whether user satisfaction alone can be a basis for comparing union catalog projects, and I propose several indicators that could be used to compare, in a quantitative way, different union catalog projects. Many of these

---

<sup>2</sup> See Association of Research Libraries websites:

<http://www.arl.org/libqual/pubs/index.html>, and Rowena Cullen, “Perspectives on User Satisfaction Surveys.” *Library Trends* 49/4 (Spring 2001): 662–686.

<sup>3</sup> Steve Hiller, “Assessing User Needs, Satisfaction, and Library Performance at the University of Washington Libraries,” *Library Trends* 49/4 (Spring 2001): 605–625.

considerations could, of course, be applied to all kinds of projects, and not only library and union catalog projects.

The main purpose of this paper is to turn the attention of the designers and coordinators of union catalog projects to the complex problems of user satisfaction and establishing measurable indicators of union catalog performance and success. The paper should be treated as a starting-point for broad discussions of the problem of assessing the results of union catalog projects with respect to user needs and satisfaction, and by no means pretends to be complete and comprehensive.

## 2 Elements of Project Evaluation

When starting a new union catalog project, the designers and project coordinators usually define its goals and the methods for achieving them in the most efficient way. They create the organizational and technical structures for the stipulated tasks, and design the timeframe for the subsequent steps. But complex projects involving many libraries, such as union catalog projects, especially in East European countries where it is very difficult to find permanent sources for financing such projects, are seldom concerned about the future results of the project in terms of user needs and satisfaction. Responsible authorities usually concentrate on launching the project as soon as possible after the funds have become available, and nobody cares about making time-consuming, and sometimes expensive, surveys of users' needs prior to defining the project goals and the project methodology. Decisions about the model, purposes, and functioning of the future union catalog are taken in small groups of project initiators and coordinators, sometimes after consultations with a few chosen librarians. How, then, is it possible to assess the project results? What actions can be undertaken to check whether the project has been a success? How can one evaluate the project and compare it with another, similar one?

There are several expressions, closely interconnected with one another, which come to mind on such occasions: project success, user satisfaction, service quality, performance indicators. Each of these terms may be the basis for considering further the assessment of union catalog project results.

User satisfaction is considered to be one of the performance indicators for a particular service. Most customer services are constantly trying to maximize the value of 'user satisfaction' indicators because it is the principal precondition for satisfying the market. But the term 'user satisfaction,' which appears to be obvious and understandable, rapidly reveals its complexity. The definition formulated on the basis of marketing considerations<sup>4</sup> is the following: 'user satisfaction' "is the emotional reaction to a specific transaction or service encounter." Moreover, apart from an emotional element, satisfaction also contains a cognitive element.<sup>5</sup> User satisfaction derived from a single transaction is determined by many different factors, including service quality, the user's past experience with the service provider, the emotional state of the user, etc. There is a close relation between user satisfaction and user needs. Users' needs are in turn shaped by historic, socio-economic, cultural and professional factors. Users in different countries, or even different user groups in the same library, may have different needs and expectations, and therefore different level of satisfaction from the same service. Because of this relative perception of satisfaction, projects that aim at providing library services such as a union catalog, and for which the measure of success is user satisfaction, should always target well-defined groups of users. The expectations of students regarding the union catalog will be completely different from the needs of librarians. Projects that would satisfy librarians would not necessarily satisfy students or researchers in our universities. Similarly, the model of a Polish union catalog might not satisfy users in South Africa, though it might satisfy Polish users' needs.

Unfortunately, there is little knowledge among union catalog designers about the concept of user satisfaction and its relation to a variety of factors including user needs or library service quality. It is commonsense,

---

<sup>4</sup> P. Hernon and E. Altman, *Service Quality in Academic Libraries* (Norwood, NJ: Ablex, 1996).

<sup>5</sup> A. Falkowski, Professor of Marketing Psychology, University of Łódź, Poland; private communication.

confirmed by scientific research,<sup>6</sup> that the better the quality of service, the higher the user satisfaction. At the same time, the term 'quality' does not need to be sharply defined. In the SERVQUAL model used by Herson and Altman<sup>7</sup> and in the work of other researchers examining service quality in the field of library and information services, quality is defined as 'perceived quality' rather than 'objective quality'. That is, it is dependent on the customers' perception of what they can expect from a service and what they believe they have received, rather than on any 'objective' standard as determined by a professional group or in conventional performance measurement.<sup>8</sup> The SERVQUAL model permitted the definition of the gaps between customer expectations and perceptions as follows:

1. The discrepancy between customers' expectations and management's perception of these expectations;
2. The discrepancy between management's perception of customers' expectations and service quality expectations;
3. The discrepancy between service quality specifications and actual service delivery;
4. The discrepancy between actual service delivery and what is communicated to customers about it; and
5. The discrepancy between customers' expected service and perception of service delivered.

Research on the boundaries of library information, psychology, and management also proved that user satisfaction may involve long-term as well as short-term perceptions, and a personal reaction to service built up over a number of transactions and experiences of varying quality.<sup>9</sup>

---

<sup>6</sup> K. Elliott, "A comparison of alternative measures of service quality," *Journal of Customer Service in Marketing and Management*, Vol. 1 (1) (1995): 35.

<sup>7</sup> Herson and Altman.

<sup>8</sup> Cullen.

<sup>9</sup> Hiller.

Should the highest level of user satisfaction also be a goal for library services, including union catalogs? According to Cullen and other researchers, definitely yes! As Cullen states,

Retaining and growing their [libraries'] customer base and focusing more energy on meeting their customers' expectations is the only way for academic libraries to survive in this volatile competitive environment.

Therefore, even though it may already be very late for some union catalog projects in East European countries, I would strongly suggest that surveys on users' expectations and needs concerning the union catalog should be prepared and carried out. Perhaps there is still time to amend or correct already decided models and schemes of cooperation.<sup>10</sup>

The variety of factors influencing actual or average users' satisfaction and their user dependence are reasons for the fact that measuring user satisfaction is mostly accomplished with direct questions about users' feelings. Questionnaires are applied to different user groups of a specific service<sup>11</sup>. Results of such user satisfaction surveys can only tell us how much a specific group of users is satisfied with a specific service. Could such results be a yardstick for comparing different projects? In terms of users' satisfaction with the project, the answer is yes, but in terms of objective performance and success indicators probably not. In the case of different union catalog projects, it is almost impossible to compare projects on the basis of user satisfaction alone, even if it were measured, because a higher level of user satisfaction from project A than from project B would not prove that project A was better, showed better service quality, was more cost effective or was used more than project B. What it would show, however, is that users of project A like the services of A more than users of project B like the services of B. Besides, one must be very careful when

---

<sup>10</sup> For example, in the Polish NUKat project, the strong focus on authority control was not counterweighed by user expectation surveys, or even by broad consultations with academic librarians across the country.

<sup>11</sup> In marketing, these 'user groups' with different attitudes towards the service or product are called 'consumers in different market segments.'

using the results of user satisfaction surveys to estimate the success of any project, because by definition these surveys are directed at the actual beneficiaries of the project and tell us nothing about the feelings of those who could or should benefit but for some reason did not. Therefore, in the case of union catalog projects, it is very important to implement user expectation surveys as widely as possible among the potential users of the catalog, and not only among the narrow group of initiators or actual beneficiaries of the project.

Despite its limited use for comparing different union catalog projects, it is still worthwhile to prepare surveys of user satisfaction, either separately by each project management or—and this would certainly exhibit good will toward international cooperation—by an international group consisting of representatives of the relevant projects, in order to ensure the homogeneity of research across different projects. The results could be used for assessing the results of individual projects and their evolution in time. The work, however, needs careful planning, and should involve not only librarians but also specialists in marketing and psychology, to ensure proper quality and methodology.

If user satisfaction is not a satisfactory indicator for project evaluation, what are the other choices? It seems worthwhile to examine whether project success might be a basis for setting up comparable indicators for the evaluation of different union catalog projects.

### 3 Project Success

A union catalog (like any other new library service), its quality and subsequent use are outcomes of the successful implementation of the project. The traditional success criteria for project implementation are based on whether the project was completed according to specifications, within the budget and in time. This very narrow view has been unable to ensure the success of an individual project. The Wideman Comparative

Glossary of Common Project Management Terms<sup>12</sup> describes user (or “stakeholder”) satisfaction in the following way:

The measure of satisfaction with project results on the part of stakeholders is a measure of project success. Satisfaction is subjective, tends to vary with time and hence is difficult to measure effectively. Project success is achieved when a project has been completed according to all requirements and satisfies the project’s Key Success Indicators.

Key Success Indicators are those project management indicators that

- are determined at the beginning of the project and listed in order of priority
- reflect directly on the key objectives of the project, and
- provide the basis for project management trade-off decisions during the course of the project

and, after completion of the project:

- are most likely to result in acceptance of the project and its product by the project’s stakeholders as being ‘successful’ in terms of ‘customer’ satisfaction, and
- can be measured in some way, at some time, on some scale.

It seems that for most union catalog projects (not only in Central and East European countries), designers and project managers have not defined any measurable ‘key success indicators’ at the beginning of the project. Even after completion of the project (i.e. after the phase of implementation) one can hardly find in the literature<sup>13</sup> any measured indicators proving that the project was really successful. After the structure has been put in place, and even after the goals have been achieved, it is too early to report, as some authors do, that a union catalog or shared cataloging project has been successful.

Before I propose several ‘key success indicators’ for union catalog projects, let us examine what the factors influencing the project and its success are:

---

<sup>12</sup> See <http://www.pmforum.org/library/glossary>.

<sup>13</sup> See <http://www.pmforum.org/library/glossary>.

## Project success factors

The literature on project implementation<sup>14</sup> identifies several general factors that determine the success of a project. The most important of them are:

1. Project mission—were the goals clear at the outset, and was there a strong sense of direction?
2. Support from top management—was management willing and able to bring to bear the necessary resources, authority and influence?
3. Project planning—was a detailed specification and schedule of activity steps produced for project implementation?
4. Client involvement—was there adequate communication, consultation and active listening with respect to all elements of the 'client system' (including the user, the stakeholder and the project champion)?
5. Personnel—were the necessary personnel for the project recruited, selected and appropriately trained?
6. Technical activities—was the required technology and expertise available to accomplish specific technical tasks?
7. Client acceptance—was the final project 'sold' effectively to the ultimate end-users?
8. Monitoring and feedback—was there timely provision of comprehensive control information at each stage of the implementation?
9. Communication—was there an appropriate network for circulating all necessary information among all the key players in the project implementation?
10. Troubleshooting—was there an ability to handle unexpected crises and deviations from plan?

---

<sup>14</sup> J. K. Pinto, *Project Implementation, a Determination of Its Critical Success Factors, Moderators, and Their Relative Importance Across Stages in the Project Life Cycle*, Ph.D. dissertation (Pittsburgh, PA: University of Pittsburgh, 1986), J. K. Pinto and D. P. Slevin. "Critical Success Factors in Successful Project Implementation." *IEEE Transactions on Engineering Management*, 34/1 (1987).

Based on the experiences of the Universe project, which aimed at the creation of a large-scale virtual union catalog,<sup>15</sup> it is possible to divide success factors for a technology-related library project, and especially for union catalog or shared cataloging projects, into three groups:

### Project factors

which reflect the overall way the project is managed and the project's information policy. Illustrative project aspects are:

- Compliance with work plan (adherence to plan, ongoing review, project management etc.);
- Visibility and dissemination (publicity for the project, raising awareness of the project, dissemination methods, Web presence, partners' involvement);
- Exploitation plans (clear action plans for partners, solving intellectual property rights problems); and
- Partner role and motivation (collaborative approach, proactive management, proper communication between project management and partners).

### Technical factors

which are related to the technical side of the project including hardware, software and maintenance. The group includes the following factors:

- Scalability (technical ability to accommodate new partners, single and stable entry point to project results, quality of service, performance, functionality, accessibility);
- Service components (application scenarios for planned services, data homogeneity, use of standards);
- Software potential (functional scope of purchased software, 'fitness for purpose'); and
- Failures and futures (servicing, maintenance, development).

---

<sup>15</sup> See <http://www.fdgroupp.co.uk/research/universe/>.

## User service factors

which are the most important from the end-user point of view:

- Integration with legacy systems and practices (use of legacy hardware and software systems, respect for best practices existing in libraries);
- Delivery of services (real user requirements, meaningful feedback from users, information resources for users, sustainable services);
- Large-scale take-up (number and quality of partners); and
- Usability (transparency of services to end-user, efficiency, flexibility).

Of course, different success factors have differing importance in different projects. For example, some union catalog projects would exhibit no technical problems, because they are based on libraries with the same library automation system. For some other project, the general factors would have less importance to success because the project has the full support of the authorities on the local (and/or national) level and is coordinated by strong, experienced institutions and people with good management skills.

## 4 Performance Indicators

In parallel with key success indicators for the project, we could define performance indicators for union catalog service. The two terms 'key success indicators' and 'performance indicators' describe in practice similar, or sometimes even the same, set of values, since the meaning of a certain measured indicator may be different and depend on the purpose of measurement. A high value for a certain performance indicator may be proof of project success. For the purposes of further discussion, I assume that all indicators proposed below are equally 'service performance' and 'project success' indicators.

Performance indicators have been defined for a classic library environment for a long time. A set of basic Library Performance Indicators is defined by the ISO 11620 international standard. In recent years, as a result of the flood of electronic services in libraries, there have been attempts to enhance

and complement the standard set with the indicators related to library electronic services.<sup>16</sup> One of such projects was EQUINOX – the project under the Telematics for Libraries Program of the European Commission.<sup>17</sup> The project lists 14 performance indicators to be used in the electronic library environment. However, only a few of them could be applied to union catalogs.

Before we discuss candidate indicators for success/performance that are specific to union catalog projects, it should be noted that the differences among the projects make it quite difficult to define these indicators. The differences among projects arise in almost all their aspects:

- The time of launching the catalog—projects are started at different times, hence it is difficult to compare them as of a given date;
- The size of the project—projects may involve many libraries, but the number of potential participants is different in different countries;
- The size of participants—member libraries are not of comparable size: some projects may involve small, specialized libraries, some big university libraries;
- Level of technology—participating libraries are at different stages of automation;
- Objectives and goals—projects have different objectives: some concentrate on providing information for users, some on minimizing cataloging cost;
- Library automation systems—projects may be homogenous or heterogeneous as to library automation systems used in participating institutions;and
- The range of the project—projects have different numbers of potential end-users.

---

<sup>16</sup> See <http://www.notredame.ac.jp/~peterson/URL/ais/standards.html>.

<sup>17</sup> See <http://equinox.dcu.ie/index.html>.

## 5 Proposed Success/Performance Indicators for Union Catalogs

Even if projects differ from each other, the indicators should not pick up these differences, otherwise indicator values would not be comparable. Also, projects should not be compared at different phases of realization. I assume that the project success indicators proposed below would be applied to projects considered to be completed. If the project is in the initial phase or less advanced in comparison with another project, it may either be compared with other projects that are in the same phase, or with earlier phases of projects now completed.

It should also be noted that all indicators may be used to study the development of a single project through time. Indicators calculated at one point of time may be compared with the values collected at regular time intervals to check whether the project is moving in the right direction; whether it is growing, or has achieved a stable phase (saturation) or has even retrogressed.

The following measures may be considered as possible indicators for union catalog project evaluation.

### The percentage of target libraries reached by the project

Every union catalog project is targeted at a certain group of libraries. It is seldom the case that an 'all or none' rule is to be applied to project members. Therefore, even with a set of project initiating libraries, there is usually some concern about how many libraries may ultimately subscribe to the project. Which of the possible libraries will do so? If project rules allow for participation of academic libraries, the potential target is the complete set of academic libraries in the country. If, for any reason, only 20% of these take part in the project, the value of this indicator would be rather low. In case of projects where agreed standards are high (and not many target libraries are able to meet them) or the project is not likely to adopt a variety of library systems, the indicator value will remain low for a long time. But this should only be a signal for project managers that the adopted design of the project was not really targeted at as many libraries as it should be. This indicator is directly related to the scale of take-up as a project success factor, but indirectly also to such factors as partner

motivation, ability to accommodate new partners, information policy and publicity, or respect for legacy systems and practices.

### Number of services in operation

Union catalog projects usually aim at more than one goal. The basic one is, of course, providing information about location and (possibly) availability of library material in a group of libraries. In case of countries where there is no central source of authority and bibliographic records (whether in a national library or a commercial institution), union catalogs try to fill the gap and, apart from providing holdings information, they aim to serve as a source of bibliographic and authority records ready to be downloaded to local library catalogs. Another goal may be assistance to inter-library loan services or support for collection management in a group of libraries. Besides, contemporary library catalog software has more and more new features that were not available before, but which are requested and appreciated by users. Examples are images of book covers, tables of contents, links to full texts, etc. An indicator value would simply be the number of different services offered by the project to end-users, although for the purposes of specific research, the set of such services must be clearly defined. This indicator is related to such success factors as service components, software potential or scope of the project.

### Number of searches per user

This indicator would be a reflection of usability and accessibility of project results. While the number of searches should be easily be ascertainable, the number of 'users' is more problematic. If the 'user' is a participating library, then the indicator would give the average number of (monthly, yearly) searches per library. Therefore, it would have different meanings for a project with many small libraries, and one in which the participants are fewer in number but are the bigger libraries. It would be much better to define 'users' as staff and registered users of all participating libraries.

### Service cost per search

This is the first of a series of proposed 'project economy' factors, and is obtained by dividing 'yearly project costs' by 'number of searches per year.' When estimating the project costs per year, I suggest that one should include only running project costs incurred by project coordinators, and not include costs accruing to participating libraries. The reason for this is that under normal working conditions, participating libraries should not incur costs related directly to the operation of the union catalog. Cataloging a new item, the bibliographic description of which cannot be found in the union catalog, has to be done anyway, whether the union catalog exists or not. Of course, all project participants have to cover the costs of the initial preparation for participation in the project: training in the new workflow, and possibly in the new software or hardware. But the body that runs the project (institutional project coordinator) has to cover many more costs related to the purchase of hardware and software, acquiring and training new staff, etc. For the purpose of the 'service cost per search' indicator, I would suggest that one leave out all kinds of initial costs related to starting the union catalog.

### Costs per record downloaded, costs per record uploaded

These two indicators are relevant for the shared cataloging part of union catalog projects. They give a picture of how expensive the project is per 'records turnover' per unit of time (year). As in the case of the previous indicator, 'service costs' should be the running costs of the project coordinator. The project would be more cost-effective (and hence more successful) if the costs per record were low. An additional indicator would be 'service costs per record in the database,' but the absolute number of records in the database (unlike the growth value) would depend very much on the phase of the project, and different projects could not be compared this basis.

Other indicators that seem to be somewhat more project-dependent, are as follows:

### Number of hits per search

This is an indicator that could show how useful the database is for users, i.e. how the database content matches users' expectations and needs.

### Number of staff per size of the database

Every union catalog project involves a certain number of people. Sometimes they are employed in the unit or institution responsible for project realization, and sometimes they are affiliated with the project only in the long run. The number of project staff depends on the size of the project and the project goals, and therefore cannot be used directly to compare projects. But if we divide the number of project staff by the number of records in the database, we would get some kind of project 'staff efficiency indicator', which may be used as an additional indicator of project cost effectiveness.

### Percentage growth of the database per year

This measure gives a picture of project dynamics and efficiency. However, it may not be constant throughout the period of project realization. In early phases, it may reflect acquisitions and retroconversion, in later phases only annual acquisitions. The measure is particularly useful for a single project and its dynamic changes.

### Percentage of expected database size

This indicator is definitely related to the phase of the project. While the number of records in the database is a known number, it may be hard to find out how many different titles there are in all the libraries participating in the project (in other words, what the target number of records is) in order to get the value of this indicator. But when calculated, it would serve as an indicator of project progress and might be used to compensate for the differences in project duration.

## 6 Conclusions

Evaluations of union catalog projects and their results should include surveys of user satisfaction and estimates of the values of a series of project success/performance indicators, defined as early as possible, even at the phase of designing the union catalog.

1. To assure large-scale participation in a union catalog project, it is highly advisable to carry out a survey of potential users' needs, taking into account predicted types and size of user groups.
2. User satisfaction is a very complex concept, and authors of user satisfaction surveys concerning union catalogs should be aware of the complicated nature of the possible results. Surveys should be as precise as possible and should be prepared with the cooperation of psychologists and marketing specialists.
3. Because of the relative and subjective nature of user satisfaction (depending on users), it is not a good or objective indicator of union catalog project success. Other measurable indicators should be defined. Examples of such indicators are given above.
4. The indicators of union catalog project success or performance should be defined as early as possible, and estimation of these values should be carried out regularly to monitor the progress of the project.
5. It is never too late to adjust the project model to achieve better user satisfaction and better values of project success indicators.



## Chapter 12

# Union Catalogs for Poets

Henryk Hollender

Je ne sais pas de lecture plus facile, plus attrayante, plus douce que celle d'un catalogue.<sup>1</sup>

While terminology relating to those catalogs that 'centralize' libraries, i.e. encompass collections of various institutions or merely physically distributed collections, is far from uniform, life will go on and not wait for a unique name. In the near future, most catalogs that are actually consulted will have to be labeled union catalogs, joint catalogs, consortial catalogs or central catalogs, catalogs *collectif* or *gesamt*, and advanced users will not even know that other types might exist. Those users will search catalogs expecting them to provide guidance through vast holdings, electronic or otherwise, not only because offering seamless passages between collections will become a standard, but also because there will actually be no libraries: there will be fabulous buildings on the one hand, and business-like organizations responsible for transmission of knowledge on the other. Organizations will need a headquarters, and public buildings will serve communities and visitors. No user will care whether the building called 'library' at the market square or shopping mall belongs to the same corporate entity as the library that she uses next to her dorm, and would be puzzled if the OPAC offered only locations like 'second floor', 'closed stacks' or 'special collections'.

---

<sup>1</sup> Anatole France, *Le Crime de Sylvestre Bonnard, Membre de l'Institut*

When we design union catalogs, we want to address as broad an audience as possible. We agree that they will be as immune to learning bibliography as poets and arts students were once considered immune to learning physics,<sup>2</sup> but at the same time we do want them to have an impact on the country's cultural policy, and we think that they have to derive knowledge and joy from catalog searches. We want to reach school students and people of divergent lifestyles.<sup>3</sup> We want to provide a common foundation for multidisciplinary studies. For some of our users, the union catalog will be the summa of the nation's culture, while the others, not necessarily able to write creatively, will read it as a text and not as a finding tool. The national union catalog in particular, once well introduced into schools, libraries, and homes, will evoke a number of disputes, and will be analyzed from religious, political, or scientific points of view. The quality of the union catalog has to be unique, as there is much more at stake in designing it than there is with the catalog of any single library.

The quality of union catalogs conceived and operating in such settings seems to depend on two sets of conditions. The first is that merging catalogs changes their scope, and the scope is not neutral but has a value; we are adding to it by introducing changes, and the better the changes are controlled, the more substantial the addition is. Any information resource has its contents and its community of expected users, and digital libraries do not seem to change much in this respect.<sup>4</sup> Technically, however, it is easier to include new resources in an electronic file, to merge files, or to augment a file rapidly than it is with printed works; contrary to any printed

---

<sup>2</sup> R. H. March, *Physics for Poets* (New York: McGraw-Hill, 1970).

<sup>3</sup> See <http://www.bertelsmann-stiftung.de/documents/ACFs3Rv.89.pdf> for information on a recent conference which covered the area of 'lifestyle' in library users and helped launch a Bertelsmann Foundation project in Eastern Europe.

<sup>4</sup> While it is only this author who can assume responsibility of the ideas and terms deployed in the paper, some of them might have been inspired by the book by Christine L. Borgman, *From Gutenberg to the Global Information Infrastructure: Access to Information in the Networked World* (Cambridge, MA: The MIT Press, 2001). On contents and communities of digital libraries, see 123–129.

multivolume national union catalog, neither the compiler nor the end-user can visually and conceptually encompass the contents in one viewing. When we search an online file, we do not find out quickly what its character is like, for instance whether it is sufficiently scholarly (or popular or educational) for our purpose—an uncertainty we would be saved when dealing with a printed work. This invites a reduction in quality: if the inconsistency of a file can go unnoticed, why care?

The second set of conditions pertains to the user's skill and cognitive style. Users who are not advanced are backward, and the level of backwardness may vary. In my educational environment, a card catalog will long remain the mother of all catalogs. Some users on the premises of the library do not turn to the online catalog at all, and they do not want to hear that they are thus isolating themselves from the only currently updated finding tool we provide. And the card catalogs are different. The future of catalogs as tools for encompassing collections of more than one library is actually the opposite of what we currently see in Poland: there are more catalogs than libraries! In research libraries that still have no online catalog, or in which the online catalog is just a special addition to the set of manual finding tools, users and librarians have a predominant peculiar feeling that the material covered by the catalog should all be of one kind; if it cannot, it is better to maintain more catalogs, one for each type of material. Thus the card catalogs seem to divide, not unite, and this because they are organized according to the habits of users who cannot tolerate the fact that divergent materials may be covered with one uniform finding tool.

Of course, this attitude emerges from the very essential link between the contents and the access points. While we do not know whether library history has explored these issues, we can guess that psychologically, cards in a catalog are equivalents to the title pages of actual books. There is more sense of order in browsing similar title pages than in browsing title pages that do not match. If the collection is well-rounded, the card catalog is correspondingly well-rounded. One collection, one catalog. Material that is foreign to one collection makes another collection, and the second collection requires a separate catalog. One does not put into a single catalog materials that do not belong together, and one hardly even cross-references them. What we see around us is a tacit consensus that having a set of catalogs in a library is normal. Even tools that are currently used and

growing in importance carry this historic badge of compartmentalization, such as the Library of Congress Classification, which is more like a set of classifications, related rather than united. When we try to win support for union catalogs, we have to take into account that for those who still prefer to divide rather than to join, projects like NUKat are simply ugly. And the argument concerning the superiority of local files over union files that we have experienced in Poland might also have this hidden cultural dimension.

The library culture in which I grew up provides some examples of how ill tolerated modern catalogs might be, no matter in what format. For instance, librarians of languages represented at the library of the Institute of Iberian Studies, Warsaw University, requested separate catalogs, and the librarian I met there, herself a scholar, saw to it that Spanish and Portuguese files remained separate forever. And when we interfiled our Russian and non-Russian serials in one card catalog at the Warsaw University Library, some users rose up with fierce objections. But it was not the feeling for the Russian language that fueled the argument. As all the Russian headings were transliterated,<sup>5</sup> there was no need to know Russian in order to use the old 'divided' catalog, and no specific satisfaction from using it for those who were proficient in Russian. I really do not know whether our reference staff used to respond regularly with the information that in the online catalog, things have gone much further: serial titles were merged with other materials' titles into one index. But there was and still is a delay in recataloging all the serials into the online format, so users kept practicing their searching skills and styles on the manual tool.

What would make sense here would be some in-depth research on how the amalgamated online catalog felt in the hands of those raised on the following set of tools: one alphabetical card catalog for books (author searches only, no title searches), one subject catalog for books, the alphabetical catalog for Russian-language serials and the alphabetical catalog for non-Russian-language serials. In some cases, preserving strict separation among files may be a pragmatic solution. We can think of an impressive example of a union catalog in which a sophisticated mechanism

---

<sup>5</sup> It happened after Warsaw University, Russian from about 1870, had switched back in 1915 to Polish as the language of instruction.

achieves interfiling of languages and alphabets: this is ULI, the Israel Union Catalog.<sup>6</sup> Nevertheless, the resulting display may be unclear for the user who does not know Hebrew. It has to be admitted, however, that ULI must have been designed for a 'bilingual' user. We have to pay close attention to solutions adopted in numerous countries of the world in which libraries are full of literary and scholarly collections in one language and alphabet, while the current publishing output is mostly in another language and alphabet.

Generally, however, the very notion of a union catalog tends to endanger the time-honored feeling of order that prohibits catalogs from becoming too wide in scope. We have to respect this feeling and take it into account in our planning, in our public relations, and in our display design, as any textual habit should be respected by librarians who are serious toward their audience. While not necessarily poets, most of our users are, and will long remain, people with a background in the humanities, whose attachment to information-seeking behavior, once acquired, is stronger than in scientists. It also has to be admitted that OPACs in Poland have long contained only traditional 'bibliothecal' types of materials, and until recently no electronic publications, nor music, maps, video recordings or microforms, and few journal titles. Even if they were in common use in most libraries, even if they contained the retrospectively converted material, those 'books-only' OPACs would help to educate rather conservative users, who might sense discomfort when exposed to more diversified contents.

By way of an invitation to explore this issue, we have to ask ourselves whether the multi-contents type of OPAC, which we as librarians want—for we naturally do want a big file—is really as easy to search as an OPAC that contains only a specific type of materials. It seems obvious that inclusion of various formats, genres, or provenances does not make it more difficult to search, because the search results depend directly only on the catalog's functionalities. But indirectly? Let us examine a case. Our (Warsaw University Library) current OPAC, supported by VTLS99, does not permit searching journal titles only. Never mind; a user that looks for a

---

<sup>6</sup> See <http://libnet1.ac.il/~libnet/uli/uli.htm>, last consulted September 2, 2002.

journal title and does not care about monograph titles or video recording titles uses the title search option and locates the journal required, or not. As a response, most systems will probably generate a title index, and if there is no hit, there is at least a good proximity-based orientation among similar titles. If there is no hit, some systems point to a place in the index, saying something like “your [searched word or phrase] title would be here”. If the search result is a list, some catalogs inform the user as to which of the titles retrieved are actually serials. What happens next is some kind of manual filtering; it requires discipline, but is not likely to mislead the researcher.

On the other hand, such searches must have been problematic to users, since in today’s OPACs we increasingly encounter sophisticated automated filtering functionalities. We can guess that with more and more contents in catalogs, such functionalities are in larger demand, and that their introduction was made possible by the progress in software engineering. Designed to limit the search scope, they seem to feed the natural need of OPAC users to acquire some kind of ‘clearer view.’ In NUKat,<sup>7</sup> supported by VIRTUA, we can decide on a Journal Title immediately after selecting Browse Search, so it is probable that theoretically a Russian Journal Title would also be possible (and, in some specific contexts, desired). Then, from the Keyword Search display, we can go to a selection of filters: Location, Publication Date, Nature of Contents (a list of over 20 items, but restricted to monographs and serials only), Format (almost 60 items!), Language, and Place of Publication. In some systems we are offered a choice of search for Journal Title Word, Phrase, or Journal Exact Title. While this functionality seems to belong to a wide range “of tradeoffs between recall and precision,”<sup>8</sup> it does not necessarily, as in some cases, require entering a word as the only way to get a hit (for example, when you remember the name of the institution but do not know whether it published *zhurnal* or *vestnik*—the card catalog would be helpless here). In VIRTUA, you can also search for words in journal titles, but you have to switch to Keyword

---

<sup>7</sup> NUKat, or *Narodowy Uniwersalny Katalog*, is a project in Poland described by Maria Burchard in “Union National Universal Catalog in Poland,” *Slavic and East European Information Resources* 2 (2001): 15–16. Free access to the emerging file is at <http://www.nukat.edu.pl>.

<sup>8</sup> Borgman 122.

Search first, which I like better, since I do not think an inexperienced user will avoid jamming the searches if they can all be initialized from the first display.

All in all, in online union catalogs, as in all modern OPACs, the power to compartmentalize has somehow returned, and there is no need to consider it a clumsy vestige of the card catalog. The file can contain a number of records unthought-of in the era of the card catalogs, but on request it can also yield subsets. It can even be designed so that it meets the need not to mix materials where mixing is prohibited by some prejudice or taboo, or just by the highly focused interest of the researcher. An example of material that should preferably not be mixed is clandestine literature, published under communism in several Eastern European countries. In Poland, the printing of such literature was very intensive and grew semi-professional in the 1980s, to eventually become tolerated by the authorities at the end of the decade and legal with the fall of communism. Since then, it has been reflected in several bibliographies and exhibits. Should it be covered by the national union catalog? It certainly should. And what do we do to make researcher notice that she is dealing with a product of an underground press? We can include notes in respective bibliographic descriptions, assigning some collective name to items, or pointing to peculiar formats, technology, or textual features of those dissident publications. Without some easy filtering, however, this literature will never be covered by a separate list, and the scholarly usefulness of such a list is obvious. We thus either have to devise adequate filtering criteria or link respective items to another bibliography, digital or not. If we fail to do so, we confirm the need for a separate publication and deprive the union catalog, no matter how rich in contents, of some of its intellectual dimensions.

Moreover, union catalogs do, and will always, offer subsets that are not fully integrated into the file. A telling example of this can be found in the Hand Press Book Database (HPB). This file, established by the Consortium of European Research Libraries (CERL) and hosted by the Research Libraries Group, Inc.,<sup>9</sup> currently contains over 1 million records from 15

---

<sup>9</sup> See <http://www.cerl.org>.

libraries (half of them *Bayerische Staatsbibliothek* records) and is searchable only by RLIN users or CERL members. The contributions from various libraries make this material of the utmost value for scholarship, yet very diversified. Editing these contributions in order to make them fully consistent, a task to be undertaken some day, will take decades. As a result, the user must keep in mind all the time the uniqueness and limitations of each file. A variety of alternative search tools, search strategies, and different characteristics of each contributor's file make navigation a task for the qualified few. The HPB labyrinth is supported by only one thesaurus, for variant place names. There is no authority control imposed over the whole file, and some of the members also contributed files lacking any authority control. The experience of my library is that with early imprints, it is arduous enough to provide authority control for authors' names; and in the file that we contributed to HPB, we had omitted the owners' and users' names—although those were recorded in our local file—to bypass the common index of personal names. We felt that in this case, authority work would make the project never-ending.<sup>10</sup>

The manual for searching the HPB has 60 pages and seems indispensable for serious searches, and reading it is a job in itself, but again only for a very competent user.<sup>11</sup> That user is given a separate set of recommendations for searching each of the files. File descriptions normally explain the cataloging practice, coverage of the file, present and absent fields, mode of cataloging, and treatment of multivolume works. A separate chapter is devoted to working with search results.<sup>12</sup> A master researcher can certainly crop the HPB to an extent hardly possible with the manual or printed file, and with comparable pleasure. Discovering individual libraries behind the aggregated material can provide some additional excitement, but it is only when all the records become really uniform that there is more room for precise and far-flung comparisons.

---

<sup>10</sup> The project, funded by The Andrew W. Mellon Foundation, is described in the current CERL materials; see footnotes 8 and 11.

<sup>11</sup> See <http://www.cerl.org/HPB/manual2.pdf>.

<sup>12</sup> Karen Coyle, "The Virtual Union Catalog," chapter 2 in this volume.

Moreover, the complexity of HPB sheds light on what national union catalogs will look like when they absorb more antiquarian material. And it is still open for discussion how one can live with the essential tension between coverage and simplicity. Wide coverage is an obvious necessity for a real national catalog, as for a national bibliography, but so is the user-friendliness of the file. A catalog that includes all publications in divergent formats, types, and languages, is much less likely to be user-friendly. But in this century we can no longer believe that the national union catalog can remain esoteric. It has a new task, unheard of in the trend-setting nineteenth century: it supports shared cataloging and must percolate through the Web; otherwise the term “information society” becomes an empty buzzword. It may be that national union catalogs of the future will have at least two versions, basic and scholarly, and only the former will be in wide use in schools and public libraries. Moreover, there will be still a demand for a printed version, impressively bound; again, it will not aggregate the material exactly the way the online version does, but will offer all the possible searches. Still, the textual habits of prospective readers of the printed edition will make them expect much more sophisticated a material than the short online version might provide.

The example of the HPB helps us return to the issue of the database coverage. The idea of creating a file that would cover the publishing output of Europe up to 1830 (in the first phase) seems the most daring project in librarianship since OCLC. In a sense, this would also end up as a world catalog. But the project is progressing step by step, absorbing a collection at a time, and the file downloaded is seldom the whole collection of early imprints in a given library. The outcome of the project is not very likely to quickly reflect the real mass of printed items in libraries, not to mention the data of those publications that have not survived. No statistics applied to HPB will have much to reveal before the file fully reflects some hypothetical complete publishing output of Europe.<sup>13</sup> The fuller the HPB is,

---

<sup>13</sup> Compare H. Hollender, “Quantification, National Heritage, and Automation Strategy,” in L. Hellinga (ed.), *The Scholar & The Database. Papers Presented on the 4 November 1999 at the CERL Conference Hosted by the Royal Library, Brussels* (London: Consortium of European Research Libraries, 2001: 38–43).

the more research opportunities it provides, but the more difficult it becomes to prepare it for successful searches. When the file reaches saturation, and search and display techniques grow with it, it may be used by very expert scholars as well as by undergraduates and sensitive members of the general public. Serving those diversified groups provides the national union catalog with the justification for the costs of having the file compiled and published on the Internet.

And in this case, we may be content to belong to a smaller nation. Poland has more of a chance of a full national bibliography and a national catalog, because it is smaller than Germany or the United Kingdom. With its printed book production from the beginning to the middle of the previous century only slightly exceeding a probable 400,000 titles,<sup>14</sup> and with current annual publishing output of about 21,000 titles of books and over 5,500 titles of serials,<sup>15</sup> we have more of a chance of a successful and complete central national database than the giant publishing countries. Of course, it does not solve the problem of what should go into the union catalog first, and what last. If we concentrate on printed materials, we may some day lose our grip over the cultural mainstream. To reflect the culture of the nation, we can think today of several types of materials and types of contents which already make a substantial contribution to the life of a country. And we can no longer stick to the sixteenth-century definition of publications. Include Web pages? Problematic, they come and go. Include printed ephemera? There has never been an adequate definition of a unit of ephemera. Include sub-cultural newsletters called fanzines? They are hopelessly local by definition, but some artistic magazines started as fanzines, and there is no fringe in culture that should be avoided by a cataloger. On the other hand, tomes of devotional literature, flooding deposit libraries in Poland today, might as well be excluded with no harm to either scholars or the general public. But when we see how difficult it is

---

<sup>14</sup> Estimation based on: M. Czarnowska. *Ilościowy rozwój polskiego ruchu wydawniczego 1501–1965* (Development of publishing output in Poland in numbers). Warsaw: Biblioteka Narodowa, 1967: passim.

<sup>15</sup> Estimates provided by the Department of Acquisitions, Warsaw University Library. WUL is a legal deposit library.

to impose standards of cataloging and compile a national catalog of books and periodicals, I think that those phenomena which are traditionally perceived as “off” will long be left for specialized bibliographies and inventories.

The politics of the national union catalog is an area we in Poland have hardly touched upon in our discussion of what NUKat should look like. NUKat is heavily oriented toward supporting shared cataloging, and in the absence of real retrospective conversion, whatever little shared cataloging the libraries will perform retrospectively will create retrospective resources. The scope of this inclusion will not be clearly understandable to the public. Libraries will merely recatalog items that they need, and those items will go into NUKat. The job may be done semi-automatically, by applying OCR procedure to old catalog cards, as we do in Warsaw with our early imprints.<sup>16</sup> Are early imprints a priority? Well, in some Polish libraries they are treasures, largely unexplored. But some other priorities could also be identified. If there is no discussion and decision, libraries will probably continue the recat by progressing backwards, and flood the file with items from the post-war period, dominated by works nobody ever wants to consult anymore. We can also focus on items requested by users, which received the fast lane in processing. Then, however, we receive a national union catalog that reflects the users’ needs but has little impact on those needs, and does not reflect the national publishing output.

The alternative policy is to select areas that are perceived as the strength of the country. In the case of Poland, this would be, for instance, mathematical logic, and indeed there is an ongoing project to get that heritage digitized; if digitized, it must also be recataloged for inclusion in NUKat.<sup>17</sup> The other area could be poetry, because Polish poets, with two Nobel Prize winners, apparently have some appeal to international

---

<sup>16</sup> Hollender 39.

<sup>17</sup> The project is coming into being on a grant from the State Research Committee for the Interdisciplinary Center of Mathematical and Computational Modelling, with the cooperation of the Warsaw University Library. See “On the Warsaw Scientific Virtual Library,” <http://vls.icm.edu.pl> (Polish version only on the date of the last consultation, September 17, 2002).

audiences. Anyway, there will certainly be choices, and there will be mistakes, too. A sample mistake is a choice of titles to be recataloged for Poland's Central File of Journal Titles.<sup>18</sup> Traditionally, Polish journals were given priority. What turned out to be missing, however, were modern, expensive international scientific journals. When recently the subscription funds became smaller and deliberations were necessary to decide which titles should be canceled, which should be kept, and which should be obtained in electronic format only, it was very inconvenient not to have them all included in our online catalog. Also, the decision taken ignored the actual habits of readers of international scientific journals, of consulting the online and not the card catalog.

Of course, union catalogs are not only national union catalogs. We do need the other types. They do exist and they will continue being published in various formats. If the information world does not set up priorities in this area and submit convincing projects to where the funding may come from, the initiative will be taken up by local historians, who will fill the gaps with their non-professional printed catalogs and bibliographies of small scope, with little chance of completion anyway.

In fact, the region of Central and Eastern Europe, due to its complicated history, should be a grateful field for union catalogs of international scope. According to an informal message from the British Library, the Incunabula Short Title Catalog may some day soon include the locations of all the libraries owning the item, thus producing the world incunabula catalog. Other ideas may be of more limited scope and cover material focused on some specific research needs. There is, for instance, an increasing interest in the geopolitical situation of Kaliningrad, the Russian enclave on the Baltic Sea. Books from dispersed German libraries in Königsberg are to be found in numerous libraries in Russia, Poland, Belarus, and several other countries. It would be beautiful to have them listed in one union catalog. Another example is the project titled Better Access to Prints from Polish-German Cultural Borderlands in the Collections of Polish Libraries, funded

---

<sup>18</sup> In the Main Library, University of Gdańsk, in cooperation with the Center for Formats and Authority Files, Warsaw University Library, currently in the process of migration into NUKat.

initially by the *Bosch Stiftung* and coordinated by the National Library of Poland. The project resulted in a number of microfilms, which can now be identified and located via a union file, available online from *Biblioteka Narodowa*, and entitled a little differently: “Early Printed Books Published Mainly Within the Territory of Silesia, Eastern Prussia and Pomerania”.<sup>19</sup> We have to admit that books printed in Silesia have always been collected by libraries in Poland, so the coverage criterion can hardly be considered clear. Moreover, the file cannot be searched by the name of the library holding the item. Nevertheless, the project provides access to some 12,000 items, is scheduled to encompass 17,000 items, and the file provides the index of search names, so it will probably be welcomed by historians. The whole undertaking really wants just one finishing touch: the step back from microfilm to the original, and the creation of a file of Germanica held by libraries in Poland. Eventually, it would have to cover items not only printed in Silesia, Eastern Prussia and Pomerania, but also elsewhere in the German language, or in historic Germany, and offer the international scholarly community access to materials from German libraries taken over by the government of Poland in 1946. I see no political reason that could prevent us Polish and German librarians from completing the first phase of such a project in 2004, to celebrate Poland’s access to the European Union as well as the 60th anniversary of the ending of World War II. Also worth considering are the prospects for union catalogs (or a catalog?) of books from Polish libraries nationalized in territories lost by Poland after 1945. Raising this issue requires some polemic with a bias among librarians, who even today voice the opinion that finding lists of any kind, when published, will support restitution claims.

With regard to increased numbers of union catalogs and their adjustment to more diversified scholarly and general audiences, which we briefly discussed under the issue of CERL and Hand Press Book Database, we will have to acknowledge, respect and influence a range of cognitive cultures. In designing catalogs—as in the case of NUKat in Poland and other possible projects—we will have to depart from our only regular customers: the university student. We will have to admit that this customer has never

---

<sup>19</sup> See <http://139.59.172.222/info/info18a.htm>.

attempted to give us a hard time: she learned quickly, worked in a hurry, asked for help if in doubt or trouble, and in most cases came to the university with some basics of computer literacy. She has mostly searched for titles or authors from reading lists. She has not understood subject searches and avoided them, thus surpassing her professors, who had hardly ever known that the subject searches existed. Yes, our student users can handle online catalogs. Still, observing the totality of our patrons, we could not help repeating after Christine Borgman: Why are online catalogs still hard to use?<sup>20</sup>

We certainly do want our union catalogs to become easier to use. To this end, we can draw on Borgman's analysis of the problem and follow her advice. That is not to say that most catalog interfaces are perfect; they should and can be improved. On the other hand, they may improve without becoming operable for everybody. This is probably a somewhat different point of view than that of Borgman.<sup>21</sup> Note that while we are increasingly depending on our skills to use sophisticated high-tech equipment, which is actually Borgman's point of departure in her most recent work, we need not make a point of making information searches literally easy. The information society will not go as far as to require identical qualities from a vacuum cleaner operator and from the author of a term paper. A catalog cannot be simplified beyond a certain point, and since what we want to retrieve is a document—a written work, a visual object, or a piece of music—we have to be ready for some textual operations. In fact, online services, permeated by conceptual, semantic, syntactic, and technical hardships, no matter how complex, will always be easier to 'read' than a medieval manuscript, because online searches are based on algorithms and after some experience we can master them, while annotations on parchment

---

<sup>20</sup> This title of Borgman's 1996 paper, published in *Journal of the American Society for Information Science*, 47(1996), 493–503, refers to her other paper, entitled "Why Are Online Catalogs Hard To Use? Lessons Learned from Information Retrieval Studies," *Journal of the American Society for Information Science* 37 (1986), 387–400.

<sup>21</sup> Christine L. Borgman. *From Gutenberg to the Global Information Infrastructure: Access to Information in the Networked World* (Cambridge, MA: MIT University Press, 2001: 116–141).

are not, and no experience can equip us against the unknown. There is no hardship if online services are based on effort, on learning, on the joy of discovering. It is all right if the acquisition of information requires some ritual, and a serious ritual is never for pleasure.

Of course, the kind of online searches that are easy have already been designed and can serve as a pattern. If we engage in Internet shopping, we are likely to be guided step by step; each step is explained without shortcuts, and once taken, it is confirmed. It might also be manageable to organize library displays the way Amazon.com is organized. We may guess that businesses generally provide easier-to-use portals because they can afford better designers, and information workers employed by libraries will always be behind in their funding and achievement. But it is our personal and intuitive opinion that designers are generally seldom good because they have a background in computing and have learned to live with texts of quite a different format and purpose than those we find in traditional finding tools: catalog cards, tables of contents, charts, and diagrams. For a patron with a long background in using a catalog with regular cards, like those required by AACR2, any screen display will seem redundant and chaotic. The display always contains some additional elements, which look as important as any other on the screen, but in fact open only some secondary option or provide some secondary message. The title page of the book, the layout of a bibliography took shape decades after the invention of printing, and we have to wait until the electronic information enters the same age of maturity.

If we again draw on personal experience, we have to admit that most catalogers and format experts, no matter how proficient, are rather insensitive to display issues, and that most Web page designers are computing experts with little understanding of a printed book layout. Indeed, in the work of many of those two groups, centuries of book design are immediately lost! Also, this has been an area in which there has been little feedback from the public. Moreover, the designer of OPAC displays has a more difficult a job to do than a person who is responsible for an ordering routine in an Internet store. There are hardly any subject searches in Internet shopping (those in Internet bookstores are on a very trivial level). And last, there are security reasons involved in commodity searches and ordering—money, credit card numbers, etc.—so avoiding mistakes is

and will always be more crucial than in information searches. It is, and perhaps always will be, a psychological issue: a person searching a union catalog will not want to devote as much time to a single search as a person deciding on an item which has to be paid for.<sup>22</sup>

However, when we look at a modern OPAC with the eye not necessarily of a poet, but of a sensitive and literate person, we have to understand why such a person feels at a loss so often. Without our own in-depth study, and in the conviction that it will be very difficult to add much to the analysis by Borgman,<sup>23</sup> we can only mention that, for instance:

- Diacritics are seldom transmitted properly, which will have an especially bad impact in catalogs containing multilingual material;<sup>24</sup>
- The lack of authority control generates noise which can be ignored in small files, but paralyzes searches in big ones;<sup>25</sup> and
- Boolean operators seem to remain a tool for the brave few, and the systems designers still prefer to make those in need activate a help screen or to turn to some help desk, which actually few will do, while only some displays show adequate advice on the same screen in which a query is to be entered or in which the search results appear.

Another problem is with subject searches. With card catalogs and with the first generation of online catalogs, it was obvious that subject searches would be avoided by most users. Those who have learned how to enter a

---

<sup>22</sup> As a result of some convergence process, however, a subset of selected items is called a ‘cart’ in Chameleon iPortal in Virtua.

<sup>23</sup> See note 18.

<sup>24</sup> If we are looking for a bad experience, we are sure to have one in searching online, for instance, for the capital of Ukraine. While the Russian version of the name of the city, Kiev, transmits well between systems and networks, the Ukrainian version as a rule comes to us with some unwanted symbols unless it is simplified as Kyiv.

<sup>25</sup> I cannot understand why a search for, say, the Ukrainian poet Oksana Zabuzhko in my otherwise favorite OhioLINK Central Catalog has to end up with the foolish advice “Your entry zabuzhko oksana would be here—Change search to oksana, zabuzko”, while in the line above Ms. Zabuzhko is properly listed as Zabuzhko, O. S. (Oksana Stefanivna).

subject query were normally satisfied with interesting findings. In big files those findings can be really exciting. But no union catalog—and indeed, no national union catalog—will have subject headings assigned in all the records in which, according to the cataloging rules, it would be appropriate. The Israel Union Catalog, for instance, does not offer subject searches at all. And in practice, there is no chance that subject headings will soon be assigned to any but a minority of NUKat records in Poland, unless there is no retrospective material and the file grows only with current cataloging.

In the future, we are likely to replace subject indexing with advanced automated indexing techniques. But the handling of subject searches is changing at present. In the first generations of OPACs, as well as in card catalogs, we had to know or just guess the actual subject term or select it from the thesaurus. It was difficult to end the search with a hit, but it was generally semantics, and not grammar, which led to noise. Currently, in most union catalogs, the system understands ‘subject’ as a ‘subject word,’ and generates a long list of supposed hits. This new trend leads toward more hits, but also more redundancy. Few OPACs are as user-friendly as that of OhioLINK,<sup>26</sup> which lists the subject headings retrieved before directing the user to bibliographic records. This is undoubtedly a nice functionality, but it does not help much, because it leaves the researcher flooded with subject terms, in which the search term actually plays the role of a qualifier or a subdivision. To retrieve what we really want (and we mostly want a supposedly proper subject), we have to either limit the search result, or activate the search template and select the search usually named ‘exact subject.’ In some newer union catalogs, ‘exact subjects’ are hidden under some separate type of search, such as “Power Search” in the California Digital Library’s Melvyl-T.<sup>27</sup> The notion of power probably refers here to productivity, and not to precision. In COPAC there seems to be no way to activate the ‘exact subject’ type of search, and due to the file size, most of the searches, especially for proper names, produce very

---

<sup>26</sup> See <http://www.ohiolink.edu>.

<sup>27</sup> See <http://mel-t.cdlib.org>, as consulted directly before the Conference.

redundant material.<sup>28</sup> It becomes clear that the designers wanted to remake the functionality of search engines like Google, where in advanced searches we always have a choice of ‘all words’, ‘phrase’, and ‘exact match’. This way we certainly make sure that no relevant material ever remains undiscovered, even in the hands of a very novice user. But we are also likely to discourage those who understand what the subject is and will turn their backs on a service that is not acting according to the accepted terminology. In NUKat, we are hoping the issue will be solved in the localization of displays; articulate guidance of the user may require adding some words of explanation that subject searches by words are actually not subject searches, but free searches within the whole field of subject heading.

Some features of new union catalogs, however, will satisfy people with ‘bookish’ textual habits. Adding a table of contents is one example; providing notes to help in better understanding the subject and type of publication is another. While the former is a novelty, the latter has always been used in cataloging, but we have seen very little of it in automated cataloging. At least one catalog—OhioLINK—offers the functionality of retrieving authors not listed in the responsibility statement. For instance, when we are looking for a poet, we can expect to find his poems included in a collection of works of various authors.

All in all, what catalogs will be understood by those who feel like strangers in the information society, or who are just beginners? What union catalogs have to be organized to meet and augment the creativity of our poets and intellectuals? Certainly the union catalog is not a field for experiment; its advanced technologies should serve a conservative purpose. It has to show the ambition to provide a database with some predictable contents and structure. It has to have clear, transparent criteria for the inclusion of material. It has to explain as much as possible to an eager reader and provide shortcuts for those in a hurry. It has to offer some kind of contract with the user: the better you learn how to operate me, the more I may assist you in depth. It has to use authors, titles, and subjects as the

---

<sup>28</sup> See <http://www.copac.ac.uk>. Let us use Kiev as an example again: as a subject it generates 1686 records, of which no more than one-fourth would have Kiev as a subject.

basic entries, and to offer author word, title word, and subject word searches with the necessary explanation, without which the lay user will never understand what a subject is. It has to produce subsets and tolerate, again with a clear explanation, that different search methodologies may be needed for each. It has to lead to the full text as quickly as possible, by providing lavish, uncontrolled information on the item contents, or by linking to the digital object, and also by facilitating shelf searches and inter-library loan. It has to look well and read well. It has to be edited. It is a publication, or even a poem in itself.



## Chapter 13

### Aiming at the Union Catalog of Polish Libraries

#### Stage 2: From the Union Authority File to the Union Catalog

Anna Paluszkiewicz and Andrzej Padziński

On October 16–18, 1997, at the Conference on Library Automation held in Warsaw under Mellon Foundation auspices, Anna Paluszkiewicz presented a paper on the Union Authority File (*Centralna Kartoteka Hasel Wzorcowych*—CKHW)<sup>1</sup>. The project described in the paper was the first stage of building the union catalog in Poland. The National Union Catalog (*Narodowy Uniwersalny Katalog Centralny*—NUKAt) was initiated in June 10, 2002, after a long period of careful preparation. The present paper recapitulates the first stage of the preparations and the key role of CKHW. It briefly discusses the development of work on the idea of the union catalog, the principal aims of the project and the requirements for the union catalog's integrated library system. Next, it discusses the preparatory work directly preceding the start of the union catalog and the start of NUKAt. Finally, it focuses on the costs and advantages of the union catalog.

---

<sup>1</sup> Anna Paluszkiewicz, "Evolution of authority file work in Poland," in *Library Automation in Transitional Societies: Lessons from East Europe*, eds. Andrew Lass and Richard E. Quandt (New York: Oxford University Press, 2000: 224–234).

## 1 Stage 1

Political and economic transformations in Poland at the turn of the 1980s and 1990s and the rapid development of modern technologies led to significant improvements in the automation of Polish libraries. The automation of many Polish libraries became possible with financial help from foreign foundations supporting Polish scientific and cultural institutions.

Library automation is a very expensive process, and can yield significant advantages only through cooperation among libraries. In order to reduce cataloging costs and accelerate the process of building databases, we need to enable data exchange among libraries. The need for data exchange has generated a demand for an automated union catalog—the source of quality data for the libraries and an efficient searching tool for the library users, enabling them to search through the collections of Polish libraries. However, the transformation in Poland and the sudden influx of funds for the purchase of hardware and software left Polish libraries totally unprepared for the implementation of this project. Since an attempt to build the union catalog in this situation might have failed, we needed first to meet the necessary conditions for the implementation of the union catalog project. Among the most important tasks requiring immediate solutions were the establishment of connections between libraries and the computer network, the preparation of rules for creating authority files, the construction of authority files and the preparation of unified cataloging rules and librarians training.

In 1992, Warsaw University Library (*Biblioteka Uniwersytecka w Warszawie*—BUW), Jagiellonian Library (*Biblioteka Jagiellońska*—BJ), Main Library of Gdańsk University (*Biblioteka Główna Uniwersytetu Gdańskiego*—BGUG) and Main Library of Stanisław Staszic University of Mining and Metallurgy in Cracow (*Biblioteka Główna Akademii Górniczo-Hutniczej w Krakowie*—BGAGH) decided to begin a cooperation. They purchased the same integrated library system, VTLS, chose the USMARC format, and began to build CKHW. CKHW consisted of the name authority file (containing records for personal headings, corporate headings, uniform titles and series titles) and the subject authority file KABA. The KABA subject headings system is compatible with two other subject headings

systems, LCSH and RAMEAU. Apart from Polish terms, the subject authority file KABA also includes their English and French equivalents. At first, the CKHW database was built only by the libraries that used VTLS; but gradually it became of more interest to other libraries as well. Until the start of NUKat, CKHW was built cooperatively by 500 librarians from 27 libraries using various softwares (VTLS: 20 libraries, Horizon: 4, ALEPH: 2, Prolib: 1). Many libraries participated passively in the project, downloading records from CKHW or from its copy created for the libraries allied by an agreement entitled "Library with the Horizon". Initially, CKHW was administered by BUW staff. Since 1996, CKHW has been supervised by the Center for Formats and Authority Files (*Centrum Formatów i Kartotek Hasel Wzorcowych*—CFiKHW) established at BUW. Standardization of headings and the use of CKHW as the only source of authority records for the catalogs of cooperating libraries maintained the consistency of data and, as a result, an effective search and exchange of data. The whole process will also facilitate the transfer of records from the local catalogs to the NUKat catalog.

Another shared project that may be considered one of the stages in the process of building the union catalog is the Union Serials Catalog (*Centralny Katalog Czasopism*—CKTCz). CKTCz was started on the BGUG server in 1995. At first it was built only by the librarians from VTLS libraries, but later it was supported by the staff of other libraries as well. CKTCz contains over 20,000 bibliographic records for Polish and foreign serials collected by the cooperating libraries. The establishment of CKTCz has accelerated the process of cataloging serials and allowed for the unification of cataloging rules for them.

Experience gained during the work on CKHW and CKTCz has been of considerable help to staff working on the NUKat project, since we have recognized the need for unified rules and procedures. In accordance with the accepted strategy, each authority or bibliographic record is entered into the central database and only then copied to the local catalogs. The control numbers of these records guarantee their unequivocal identification. This solution allows us to restrict any required modifications of records to their versions in the central database. Then the files with modified records are transferred to the local catalogs where they replace earlier versions of the records.

## 2 Towards the Union Catalog

In 1996, at a conference in Kraków, Anna Paluszkiewicz discussed the proposal of building the union catalog for academic libraries as a facility for the retrospective conversion of catalogs<sup>2</sup>. This proposal served as a starting point for the 1996-1997 idea of a Union Catalog for Academic Libraries (*Wspólny Katalog Bibliotek Naukowych*—WuKa) designed by the staff of VTLS libraries. This catalog was intended to reduce cataloging costs and accelerate the development of the online catalogs of Polish academic libraries, as well as provide library users with the ability to carry out effective searches. WuKa was originally developed for VTLS libraries, but it also allowed cooperation with libraries using different software.

It should be stressed that the beginning of 1996 was the most suitable moment for the initiation of such a catalog. After difficult beginnings, when the creation of bibliographic records was slowed by the necessity of building the appropriate authority records, the process of cataloging gradually accelerated. However, the same book was still cataloged many times in different libraries.

In the fall of 1997, the Mellon Foundation indicated the possibility of a grant for establishing the union catalog, on the condition that the catalog was to be of national character and the group of cooperating libraries included the National Library (*Biblioteka Narodowa*—BN) and the VTLS and Horizon libraries. The work on the National Union Catalog, NUKat, project started in January 1998 and included representatives of BN, the VTLS libraries (from Gdańsk, Kraków, Lublin, Warsaw, and Wrocław) and Horizon libraries (from Łódź, Toruń, and Poznań). In December 1998, the Mellon Foundation awarded a grant of \$ 705,000 for the implementation of the project. However, further work on the project was impeded by disagreements concerning the shape of the union catalog and the methods of implementation of this project. The most contentious point was the necessity of using unified cataloging rules, and what followed from this, the necessity of using the authority file. Finally, in June 2000, under pressure

---

<sup>2</sup> Anna Paluszkiewicz. "Propozycja ułatwienia retrospektywnej konwersji katalogów współpracujących bibliotek akademickich." *Przegląd Biblioteczny*, 4 (1996): 315–322.

from the Mellon Foundation, the participants of the project reached a compromise and managed to define the main aims of NUKat, which permitted the work on implementation and initiation of the union catalog to begin.<sup>3</sup> Since then, the project has been run and supervised by CFiKHW (transformed in 2001 into NUKat Center). The compromise included, among other things, the decision to provide access to non-uniform data built without authority control by means of collecting them in a separate, temporary database. This solution was abandoned in July 2001, after the KaRo catalog (Distributed Catalog of Polish Libraries), allowing for the simultaneous search of many catalogs, had been started on the server of the Main Library of Nicholas Copernicus University in Toruń (*Biblioteka Główna Uniwersytetu Mikołaja Kopernika w Toruniu*—BG UMK).

The NUKat catalog has been expected to allow for the realization of three basic aims:

1. Constructing a source of high-quality records to be used in the local catalogs;
2. Creating a source of information on the collections of Polish academic libraries; and
3. Facilitating the process of inter-library loans.

The NUKat Coordinating Group faced the difficult task of choosing an integrated library system that would allow for the realization of the accepted project. To make sure that the decision made was appropriate, we began by defining the essential features that the software for the union catalog had to possess, following the rule that data are the most important part of every system. Since it is considerably easier to replace or update software than to modify improperly entered data, the system had to permit the entry of quality data and guarantee that they were adequately controlled. It was decided that the most important features of the system were:

---

<sup>3</sup> “Kompromisowa koncepcja Narodowego Uniwersalnego Katalogu Centralnego NUKat.” *Bibliotekarz* 2000, 10: 3–4.

1. Support of MARC21 formats for bibliographic and authority records;
2. Adequate support of links between authority records and bibliographic records and mutual links between authority records, allowing the maintainance of a proper structure of bibliographic data;
3. The possibility of entering online all types of authority records;
4. The possibility of entering authority records irrespective of bibliographic records, which was important because of the additional function of NUKat as a union authority file and the principle that the record for the authority heading had to be created prior to the bibliographic record in which this heading was to be used;
5. Adequate support and presentation of data in the authority file (generating references, displaying notes, etc.);
6. Support for the character sets employed by the cooperating libraries and the possibility of converting data to a required character set. Polish libraries use character sets: ISO 6937/2, ALA and UTF-8 in their automated catalogs;
7. The possibility of keyword searching in the authority records (important for the proper support of CKHW),
8. Protection tools against uncontrolled modifications in the database;
9. Support for three subject headings systems employed by the cooperating libraries or the libraries intending to enter into cooperation. After some discussion the participants of the project decided that NUKat had to employ
  - KABA subject headings system—mainly for the academic libraries,
  - BN subject headings system—for National Library and several public libraries,
  - MeSH—for medical libraries;
10. The possibility of entering location data and building hyperlinks between these data and the local catalogs of the cooperating libraries;
11. The possibility of displaying authority records through WWW gateways;
12. An implemented Z39.50 protocol.

None of the library systems we considered met all the requirements listed above. The analysis of various systems resulted in the choice of the Virtua Integrated Library System developed by VTLS, Inc., which met most of the conditions for entering quality data. VTLS, Inc. was required to modify the Virtua system to the specific needs of the NUKat catalog.

### 3 After Choosing Virtua ILS

The decisions concerning the choice of the system and the institution responsible for the implementation of the project were followed by work on

1. building the union set of records for extended subject headings;
2. preparing the unified rules for bibliographic records
3. the purchase of hardware;
4. negotiating the contract with VTLS, Inc., and
5. organizing the structure of NUKat Center.

The creation of the union set of records for the extended subject headings was one of the fundamental tasks. Before January 2001, the subject authority file consisted mainly of records for simple (non-extended) subject headings, subject subdivision records and reference records. The extended subject heading record was entered into CKHW in only three cases: when the heading contained non-floating subdivisions, when the heading was quoted in other records as an example, or when the creation of a record for the extended subject heading was necessary for the fulfillment of semantic relationships in the subject headings system. The remaining records for the extended subject headings were entered only into the local catalogs, and they were not controlled by CKHW. Before loading bibliographic records from the local databases into the union catalog, it was necessary to verify the extended subject headings in these records. Moreover, it was essential to unify the procedures for entering and modifying records for all types of headings. Thus the modifications of records for extended subject headings needed to be entered only in the central database and transferred, like all other modifications, to the local catalogs. As these records were not given control numbers (010 tags), the integration of data from the local catalogs into one database was made considerably easier. In January 2001, we

loaded the records for extended subject headings from CKTCz into CKHW, and from that time all new records were entered into CKHW. After we had designed all necessary programs and scripts, we also began loading records from the local catalogs. The loading procedure began with generating a file with records for the extended subject headings without control numbers from a given local catalog. Records from the file were automatically provided with control numbers after the file was taken over by CFiKHW, and loaded after this modification into the buffer. CKHW administrators checked and validated all the records (those entered online as well as those loaded in a file). At night, the system generated a file of new records for extended subject headings, which was used to replace records without control numbers in the local catalogs with records for the same headings from CKHW. As a result, we accomplished two tasks: building the union set of extended subject heading records, and enabling the identification of the records for the corresponding headings by the same control number in all catalogs. In this way, all modifications in the records for the extended subject headings can be easily transferred to the local catalogs. The loading of records for the extended subject headings was finished in March 2002. By March 31, 2002, CKHW contained 195,302 records for the extended subject headings (including 139,874 loaded in files and 55,428 entered online).

Another important problem to be solved before the start of the union catalog concerned the consistency of cataloging rules. It should be stressed that recently we have observed in Poland a significant tendency to standardize the choice and form of headings. In 1998, we received official approval of the Polish standard for personal headings (published in 2000). In 2000, the Polish Committee for Standardization approved two new standards, for corporate headings and uniform titles (published in 2001). Establishing these standards was possible due to the experience gained during the process of designing rules for the name authority file and the procedure of entering data into this file (five out of the six authors of these rules worked in the libraries using VTLIS software). Since the local catalogs copy the authority records from CKHW, headings from these catalogs are identical to those in the central database. On the other hand, we have observed many differences in the local catalogs concerning the bibliographic records, despite numerous attempts to enforce unified rules for their creation. Differences pertain mainly to the method of cataloging

multi-volume books, and the choice of headings under which the bibliographic description is entered into the catalog. To solve this problem, we had appointed a group whose task was to prepare the unified rules of cataloging. The decisions taken by this group would be of considerable help while entering new records into the union catalog. Differences in cataloging the multi-volume books make the integration of data from the local catalogs very difficult. It has been assumed that the union catalog will contain exactly one record for a given edition of a given book. Therefore the working group would have the additional task of defining the procedures for data transfer from the local catalogs into the union catalog, while protecting the latter from the input of duplicate records.

After signing the contract in March 2001, VTLS Inc. began modifying the Virtua system in accordance with the requirements defined in the Appendix to the contract. At the same time, CFiKHW initiated the purchase of a server (SUN Enterprise 450) and a database management system, ORACLE 8. On August 2, 2001, the server was delivered to the library, and on the next day the test database of Virtua was installed (Release 37).

In June 2001, CFiKHW was transformed into NUKat Center, with two new sections: the Section for the Control of Bibliographic Records (9 persons) and the Section for the Control of Extended Subject Headings Records (5 persons). The salaries for the staff of the new sections are financed by the Scientific Research Committee (*Komitet Badań Naukowych*—KBN). The first stage of work for the new sections was devoted to intensive training of new staff.

#### 4 After Installing Virtua Test Database

The Virtua release installed in August 2001 had only some of the features defined in the contract. The time necessary for preparing the final version of the system and building the set of records for the extended subject headings was devoted to training the staff and testing the system.

In August and September 2001, VTLS, Inc. representatives ran training sessions for selected staff of the NUKat Center. In October and November we prepared printed instructions in Polish. Training in November and December involved all the remaining staff of the NUKat Center. In January

2002, we began training librarians from 27 libraries entering data into CKHW. As the whole group amounted to some 500 people, the NUKat Center ran 4 sessions for 4 days each for 40 librarians who later led training in their libraries for other staff. This group later met at several one-day meetings to discuss the new features of Virtua system and the procedures for entering and modifying data in the union catalog.

After installing Virtua, we immediately started thorough testing. In the beginning, we prepared the list of the system features defined by the contract that were not implemented with the first release of the system. VTLS answered with a schedule for installing the subsequent releases of the system that would include the requested features. Before the final approval of the system, the NUKat Center tested about 10 Virtua releases and engaged in very intensive discussions with VTLS via e-mail. The fruitful cooperation between August 2001 and May 2002 saw the exchange of 600 letters.

## 5 After Final Approval of Virtua

In May 2002, we agreed on Release 40.2 of Virtua, which had all the features needed for the start of the NUKat catalog. We settled the conditions for the migration of CKHW records from the classic VTLS version to Virtua. We also defined the parameters for the NUKat database and the permissions for various groups of Virtua users. We also prepared the scripts necessary for the proper functioning of the process of entering, modifying and validating data. CKHW was migrated from May 25 to June 6, 2002. At that time the database contained 721,425 records, including 407,042 records for personal and corporate headings, 39,221 records for series titles, 6,565 records for uniform titles, 3,578 records for combined name/title headings, 60,855 records for KABA subject authority headings, and 204,164 records for extended subject headings of the KABA subject headings system.

Since June 10, 2002, all new authority records have been entered into NUKat, which also functions as the union authority file. On July 5, 2002, NUKat opened its bibliographic database for the first bibliographic records.

According to the NUKat project, the union catalog will take over the bibliographic records from the catalogs employing the CKHW headings. In September 2002, we loaded the bibliographic records from CKTCz (21,943 records). This task was considerably facilitated by the fact that CKTCz was built as a union database not allowing any duplicate records. After the transfer of CKTCz data, all new bibliographic records for serials are now entered in NUKat and the CKTCz database at the server of Gdańsk University has been closed. This operation will be followed by the more difficult transfer of bibliographic records from the local catalogs of individual libraries.

In view of the fact that NUKat replaced CKHW, we first trained the librarians from the libraries that cooperatively built this authority file. In October 2002, we started training for other libraries that entered into cooperation with NUKat.

At present, one of the more important tasks is adjusting the system to the needs of Polish users, and modifying and completing the Polish version of system and help messages.

## 6 Procedure of Entering and Downloading Data from NUKat

The NUKat catalog is supported by the Virtua system. The libraries cooperating with NUKat and using a Virtua client to enter data employ one of the following systems: Virtua, classic VTLS (gradually they will be migrating to Virtua), ALEPH, Horizon and Prolib. Virtua employs the UTF-8 (Unicode) character set and the local databases employ ALA, ISO 6937/2 and UTF-8 character sets. This difference has considerable impact on the procedure of copying data from NUKat. VTLS, Inc. provided the Virtua client and the VTLS client (EasyPAC) with tools enabling the transfer of authority records as well as bibliographic records from the Virtua database to the VTLS database. In the case of systems with an implemented Z39.50 protocol (ALEPH, Horizon), the bibliographic records may be transferred online from NUKat to the local catalogs. The modified bibliographic records are transferred to the local catalogs in files generated at night. We also prepared procedures enabling the file transfer of new records built by a given library. These procedures will apply to the libraries

that enter many records into the union catalog. Unfortunately, the Z39.50 protocol does not permit the transfer of authority records. Here we may offer two solutions. One is to save records to the local disk (using a Virtua client) and load them to the local catalog by means of the client software used by a given library. This solution is employed by ALEPH libraries. Horizon libraries copy authority records from a CKHW copy at the server of BGUMK in Toruń. Another copy of CKHW is kept at the server of the Main Library of the Silesian University (*Biblioteka Główna Uniwersytetu Śląskiego*—BGUŚ) and used by Prolib libraries. Both copies are updated with new records entered into NUKat by files generated at night. In a similar manner, data in CKHW copies and the local catalogs are updated by files generated at night containing modified authority records.

It should be stressed that the Virtua system is provided with a buffer that protects the database against any uncontrolled modifications. Every new or modified record is checked and approved by NUKat Center staff. At night, the approved records are transferred from the buffer to the database proper. On the one hand, this procedure allows us to control data entered into the union catalog, and on the other it helps to protect the consistency of NUKat data with data in the local catalogs and CKHW copies, since the files with new and modified records are also generated at night. The only data that may be added to a bibliographic record without the control procedure described above are the symbols of libraries that own the document described in the record and have copied this record to their local databases.

## 7 Presentation of NUKat Data

Apart from the usual ways of searching in online catalogs (browse search, keyword search in bibliographic records, control number search), the Virtua client provides other search methods that may be of more interest to a librarian: a search through the content of fixed fields of the bibliographic record, and a keyword search in authority records. The Virtua client is used exclusively by the librarians entering or copying data from NUKat. Other users may access the database via the Chameleon iPortal. The Chameleon iPortal provides full access to the OPAC features of the Virtua client. One of its more useful features (to Polish as well as foreign users) is the option

of displaying an authority record. The Chameleon iPortal has been configured to provide hyperlinks to local databases. This feature means that the library user viewing a bibliographic record can select a special link derived from a library symbol, jump to a given local database where the specified search is performed, and receive information on the availability of an appropriate item. This is a method of accessing the local catalogs supported by Virtua and VTLS. The Chameleon iPortal also makes it possible to search multiple databases with a single query (broadcast search). This feature will be used for broadcast search in the catalogs of non-VTLS libraries.

## 8 Our Reasons for Choosing the Strategy of a Union Catalog

The maintenance costs of a union catalog are very high. Irrespective of the expected benefits, we should strive for cost reductions. We have decided to transform CKHW into the union catalog, which has allowed us to reach two objectives: CKHW has been moved to a new server with new software (more comfortable for the librarians entering data as well as for users), and we built the union catalog. Supporting only one database is less expensive and more convenient with respect to the data entry process. Moreover, the transformation of CFiKHW into the NUKat Center helped in reducing the costs of maintenance of the union catalog to an unavoidable minimum. Present and future costs should easily be outweighed by the expected benefits. However, the six months of work since we started entering the bibliographic records into NUKat is too little to provide sufficiently credible data that would confirm the advantages of running a union catalog. But our general observations over this period seem to confirm the appropriateness of our choice. From July 5 to December 31, NUKat entered 52,687 bibliographic records (24,051 records for books published in the years 2000–2002, 27,761 records for books published before 2000, 875 records for serials). During the six months, 32,545 bibliographic records for books were copied to the catalogs of the libraries other than the library that created a given record. This number has been increasing quickly in the case of records for books published in the years 2000–2002. In the group of records for books published before 2000, only 1,043 records have been

copied. These records will be more intensively used when the subsequent libraries begin the retrospective conversion of their catalogs. We can already see the positive side of entering and copying data from only one database. Libraries copy records entered into NUKat by several different libraries. These records would not have been copied if the libraries had had to search for them in many catalogs. We want to stress the fact that before NUKat was established, records were copied from the library catalogs employing the same software. NUKat catalog allows for copying records entered by the libraries employing various systems. With the participation of former CKHW libraries, we expect NUKat to enter about 8,000–10,000 bibliographic records a month, copied on average by two local catalogs. The librarians from those libraries should be able to catalog a one-year influx of documents in about six months. These numbers will change further as new libraries enter into cooperation and passive users of the union catalog employ the results of the work of the former group. If the librarians from all libraries participating in NUKat project start entering the bibliographic records, the one-year influx of documents may be cataloged in about 4 months. The remaining time can be devoted to other tasks.

As mentioned earlier, the NUKat union catalog is built on the basis of cooperative cataloging. The procedure of creating the authority records for headings precedes the procedure of building the bibliographic records in which these headings are to be used. The cooperating libraries tend to describe each document only once, as each authority or bibliographic record is created at first in the union database and only later copied to the local catalogs. This solution has proved to be successful in the case of libraries building CKHW (as regards authority records) and CKTCz (as regards bibliographic records for serials). All cooperating libraries receive daily files with modified records, which permits the automatic transfer of union database modifications to the local catalogs of cooperating libraries. This scheme is very useful in NUKat due to the specificity of subject cataloging. Most Polish libraries begin with creating bibliographic records without subject headings (they are added later by another group of librarians). As a result, new records entered into NUKat often do not contain subject headings, and they are copied in this form to the local catalogs. After they are completed with subject headings, they are transferred again into the file with modified records. This allows for the

reduction of time-consuming subject cataloging to a single operation on each title. Before NUKat the same task of subject cataloging was performed many times by different libraries.

The above approach to a union catalog may seem expensive and time-consuming. It results from a belief that modern integrated library systems, technological progress in the network accessibility of databases and their connection by means of Z39.50 protocol will bring the expected benefits only with high-quality data. Low-quality data may undermine the value of a system, even if other elements (software, hardware) are of a very high standard. Therefore, it is of primary importance to rely on a union authority file, as well as to organize the process of data entry in a way that permits cost reductions without loss of data integrity. Data integrity will improve the broadcast search efficiency in the KaRo catalog, which is complementary to NUKat. It also should be emphasized that the solution we have accepted, avoiding duplication of work, will accelerate the process of cataloging new books in libraries. Time saved in this way can be devoted to retrospective cataloging. The whole process described above should considerably accelerate the development of automated library catalogs and improve the system of inter-library loans.



# **Chapter 14**

## **Implementing KaRo: The Distributed Catalog of Polish Libraries**

Tomasz Wolniewicz

### **1 Introduction**

This chapter was written almost exactly one year after the official launch of the Polish distributed library catalog KaRo. We discuss the functions, limitations and successes of this service, as well as problems and lessons learned for the future and some general observations that can be applied to similar distributed services. The system is under constant development, and the most important features of the new version are described at the end of this chapter.

### **2 Background**

Ever since library catalogs became accessible via the Internet, the need for a coordinated access system to bibliographic data has become apparent. In Poland, the demand for such a system arose from two main directions:

- Reference services, which help users to locate information, often leading to inter-library loans;
- Cataloging, where access to bibliographic data prepared by other libraries drastically reduces cataloging time.

Since the number of library automation systems is rather limited, libraries naturally create groups that use the same software. In Poland, in each of these groups, libraries established their own ways of cooperation for

transferring each others' records. However, things were much more complicated for libraries from different software groups, and even libraries in a single group did not have systems of distributed information service (even if it was technically possible to install such a service).

It should be noted that the views of the present author may be influenced by the fact that he works in a particular library. Nicholas Copernicus University uses the Horizon system, as do some 50 other Polish libraries. These libraries form a very differentiated group, ranging from relatively small to quite large, and from narrowly focused to completely general. This is quite different from the Polish VTLs group, which consists of large academic libraries and is traditionally a leader in standardizing library automation in Poland.

The growing pressure towards a unified service resulted in a successful grant application to The Andrew W. Mellon Foundation for the creation of the Polish Union Catalog (NUKat). The process of defining the role of NUKat turned out to be much more complicated than initially expected. The general assumption that the catalog would contain bibliographical data together with pointers to libraries was never disputed, but there were different approaches to how the data should be entered, which libraries could be represented, etc. One approach was to load the catalog quickly with data from very many libraries, in order to have a wide information service (with a lot of record duplication). A second approach, which was ultimately adopted, was to take every possible precaution against poor quality and duplicated data. The decision to take the second option meant that the widely understood informational role of the catalog will not be realized very soon, which left room for an alternative (distributed) system. Such a project, named KaRo, was launched on July 20, 2001, and is now officially seen as a complementary service to NUKat.

### 3 KaRo in Practice

The system is available on the Internet at the address <http://karo.umk.pl>, and provides access to 60 Polish library catalogs (including NUKat) and (after selecting the 'World' option) to nearly 20 additional foreign libraries. The language of KaRo can be switched to Polish or English (although help

screens for the English version were not yet complete at the time of this writing). The user can enter up to three search terms, and select libraries either individually or by predefined groups (university, technical, genera, etc.) or by simply using the 'select all' option. The user also controls the maximum length of time in which the search must be completed, the number of brief results shown on the screen and the type of display of distributed search results.

By limiting the location to one Polish city, the user can turn KaRo into a search service that can specify which library in a given city to go to.

Distributed search results are shown as a list with the number of hits in every selected catalog. In the standard view, this list is sorted into groups in which the search resulted in success, in which nothing was found and some errors appeared and in which a timeout occurred. Each entry in the list leads the user to an individual library where he or she gets access to various details. The first screen for a single library presents results in brief with several records on one screen. By selecting a record, the user is taken to the full view, which shows all relevant fields in the bibliographic record and, if the library provides this information, also holdings details. In the case of journals, the holdings are shown in two levels of detail and can be displayed in ascending or descending order. If the 856 MARC field is filled, the user can get direct access to the electronic source described in this field. In the case of journals, this is usually the link to the electronic version. In the case of the Polish ALEPH libraries, the link leads to the record in the original library OPAC, where some additional information can be found. From the full view, the user can switch to a tabular MARC view. From both full and MARC views the user can save the binary MARC record as a file. The popularity of each view is shown in Table 1

Instead of the standard list of distributed search results, the user can choose to receive the results 'as they come.' In this mode it is not necessary for the entire search to be completed, since first results are available almost immediately, and if they are sufficient, the user can move on much more quickly. The disadvantage of this approach is that the formatting is poorer and no sorting into categories is possible. This display format can be also enhanced to provide the function of sending some initial records from each library. This puts a heavier load on the individual library system, and the formatting of the result is currently rather unpolished.

**Table 1. Percentage Preferences for Views**

<b>View type</b>	
Brief view with several (default = 5) short records on one screen	55.6%
Full view in a 'user-friendly format' showing most important fields	35.6%
Full view in MARC format (all fields)	6.7%
Downloading a binary MARC record	2.1%

The initial screen of KaRo also serves as a link to libraries' home pages and to the KaRo single-library mode, in which the user uses KaRo simply as an interface to one library. This has the advantage of providing a well-known tool, rather than having to get accustomed to a new interface for each library system. Unfortunately, it turns out that only 5% of all operations are performed in this mode.

## The Users

During one year of service, KaRo has answered over 960,000 queries, by which we mean all accesses to the system that required sending bibliographic data (including a switch of format from standard to MARC). The monthly maximum equaled 124,784 queries in June 2002, with a daily maximum of 7,029 and hourly maximum of 1,349. About 20% of all queries are distributed searches, and the rest correspond to accesses to information delivered by a distributed search. This ratio seems quite stable both in short and long-term observations.

In spite of the very heavy usage, the user base of KaRo is not very large. Over 9,000 different Internet addresses have been seen, but only half of them used the system more than 10 times. The exact distribution of clients is shown in Table 2.

Among the biggest clients, three belong to one public library and in total have used KaRo nearly 90,000 times. There were 514 regular clients who used KaRo more than 50 times and were seen in 5 different months. On a typical day, between 150 and 200 different Internet addresses are observed and over 5,000 queries are answered.

**Table 2. Number of Visits and Clients**

<b>Number of visits</b>	<b>Number of clients</b>
1 – 9	4,644
10 – 49	3,033
50 – 99	633
100 – 999	728
1,000 – 9,999	146
10,000 – 49,999	11
<b>Total clients</b>	<b>9195</b>

Most accesses come from higher educational institutions, public libraries and research institutes, but there is also a significant client base on leased lines supplied by various Internet providers, many of which can be home connections. Some accesses from outside Poland are also seen, but not very often.

KaRo is quite popular among Polish librarians as a cataloging aid. Therefore it may seem surprising that the MARC view is much less popular than the ‘user friendly’ format view. One of the possible explanations is that in the current version, the user is forced to go through the standard ‘full’ view, in order to get to the MARC view, and if a new search is performed, the results will always be displayed in the brief view (even if there is only one hit). These are obvious limitations that have already been corrected in the next release under preparation. The reason for the relatively low interest in downloading binary MARC records is probably the difficulty of loading such a record into the local database, especially as the record is saved exactly as it was stored in the supplier database, possibly in a coding format different from that of the local database. Adding a planned translation service to KaRo should help to deal with this problem.

## 4 Implementation

The idea of using Z39.50 as a basis for a distributed search engine is not new. There are many examples of such systems, of which the Canadian vCuc<sup>1</sup> is probably closest to KaRo. Initiatives like Bath Profile<sup>2</sup> have been established mainly to facilitate a distributed use of Z39.50 by making individual libraries adhere to a common set of standards. There are several features making the KaRo project different from many other Z39.50 based distributed search systems:

- It is a one-man project;
- It is based entirely on free software;
- It requires only minimal cooperation from participating libraries, as all configuration differences are handled inside KaRo; and
- It keeps virtual sessions open indefinitely.

Access to library catalogs is performed via Z39.50 protocol; hence, only libraries providing Z39.50 servers can cooperate with KaRo. Unfortunately, this currently excludes several important libraries.

The core of the system is written in Perl and relies heavily on several publicly available software packages. The main Z39.50 functionality is provided by specialized packages (ZetaPerl in the current version and yaz<sup>3</sup> in new versions), which have been slightly modified. MARC record handling is done by the MARC Perl module, Unicode transliterations are done by the Unicode module, ISBN is handled by the ISBN module, and the Web interface is written with the help of the CGI module. The main user interface is written in PHP and JavaScript.

---

<sup>1</sup> vCuc—Canadian virtual catalog run by the Canadian National Library: <http://www.nlc-bnc.ca/8/6/index-e.html>.

<sup>2</sup> The Bath Profile: An International Z39.50 Specification for Library Applications and Resource Discovery. <http://www.ukoln.ac.uk/interop-focus/bath/1.1/intro.html>.

<sup>3</sup> Home page of Indexdata providing the free yaz toolkit: <http://www.indexdata.dk>.

KaRo is installed on a dedicated two-processor PC running under the Linux/RedHat system. The main program runs continuously, so that only very limited code needs to be started for every connection. This solution complicates the design, but dramatically increases performance and lowers memory consumption. From current observations, it is quite obvious that this system will easily handle a tenfold increase in connections.

Even though Z39.50 is an international standard, individual vendor implementations vary in many small, but important details. In addition, installations in libraries also vary, for instance in the handling of extended characters, the meaning of certain local MARC fields, etc. For these reasons KaRo has quite extensive configuration possibilities, where all these small details are handled. The configuration is much more extensive than in a typical Z39.50 client. Anomalies that need to be taken care of are, for instance, different character encodings in a single record, where the bibliographic part may be encoded differently from the holdings part. Configuration also controls the load, which will be described later.

There are several commercial products available that, at least in theory, can perform the functions of KaRo. Many such systems are in operation throughout the world. Still, there are some good reasons why such a system should be written from scratch:

1. There is enough free software for the realization of various parts of such a system to ensure that the programming task, while non-trivial, is not overwhelming;
2. Writing a system and using software available in source helps to solve some of the problems of closed products. There were cases in which some Z39.50 server implementations were faulty, which led to strange behavior by client software. These problems were overcome by modifying the Z39.50 tools used inside KaRo,
3. With full control over the software, new features can be added easily, but with commercial software, one is limited by the system configuration. In the earlier KaRo versions, one of the software libraries used internally by the system was distributed freely, but in a precompiled form with no access to the source. This created a problem that could not be overcome, which led to the decision to write a dedicated Z39.50 Perl module based on the `yaz`

software library. Commercial products are expensive and often have license limitations, while Polish libraries have very limited budgets.

Every Web database interface has to implement the notion of a user session. This is particularly important with Z39.50 systems, as a typical access consists of two steps, search and presentation, where presentation of records is based on information provided by the search operation. The Z39.50 server has to keep information obtained from the search operation for future presentation operations. It is obvious that neither the distributed catalog system nor the library Z39.50 server can keep a session indefinitely. It is therefore quite typical for such systems to time out and tell the user to start the session from scratch. Such behavior can be quite irritating, and KaRo produces its Web output in such a way that it can regenerate all information from the output page even if a session has been closed.

## 5 Load Control

A distributed search system can place a significant load on the resources it uses. At current usage, up to 1,300 queries per hour are serviced. Individual library receives up to 400 queries, but typically not more than 200. Even though that does not seem to be very many, some limits may have already been reached. Here are two main reasons:

- Library consortia typically use a single machine to service many databases. If this happens, the 200 per hour may grow to 2,000 or more, and what is worse, distributed searches hit the server at the same moment with queries to several databases;
- A Z39.50 session normally lasts through the whole of a user's interaction with the database. If the library has a license limit on such connections, there may be a problem both for connections from KaRo and for connections from the local system.

To make the situation less drastic, KaRo can be configured so that within one distributed search it will not send too many queries to a single machine. This lowers the load on the servers, but produces timeouts if the timeout limit is set too low. If KaRo calculates that due to timeout limits and load limits, some libraries will not be reached in time, it immediately sends the

‘timed out’ report and does not even contact the library database. Since after a distributed search the user will choose a single library from the whole list, it makes no sense to keep all connections hanging; therefore after the distributed search all connections are closed. When a user chooses an individual library, the search is run again (using the KaRo session regeneration mechanism), and this new single session is then held throughout the user interaction. This solution pays some performance penalty, but the overall performance gain and lower load on individual servers make this approach optimal. If a library has very limited license resources, the session timeout may be shortened. This will lower the performance and introduce more operations, but may be a better choice than allowing unused sessions to hang and use valuable licenses.

Currently, there is no overall load control for multiple sessions.

## 6 Lessons Learned

Running the system for one year, studying statistics and talking to many users has provided some interesting information on user behavior and preferences. We describe some below.

### Navigation

KaRo tries to help users by remembering their settings and eliminating unnecessary Z39.50 session initializations. In order to take advantage of this, users must navigate by clicking on the ‘new search’ link, visible on every page. Unfortunately, this style of navigation is used rather infrequently; it seems that users prefer to navigate using their Web interface ‘back’ button. A better way of handling users’ individual settings should be put in place.

### Multiple term searching

KaRo allows up to three search terms connected with the logical ‘AND.’ We have decided not to allow the logical ‘OR’ operation, since using several terms is mostly done to reduce the number of hits and not to widen the search. The only possible case for the ‘OR’ would be with subject searches,

where the user is not quite sure of the exact subject classification. Since there were no requests for this functionality and the user interface would have to be a little more complicated, it seemed that it was better to keep only the logical 'AND.'

About 80% of all distributed searches use a single term. 55% of these are title searches, 28% author, 11% ISBN and only 2% subject. Two-term searches account for 17.5%; 95% of these are a combination of author and title, 2% of publisher and title. Only 1.5% of all searches use three terms, half of them a combination of author, title and publisher.

Taking into account the fact that KaRo is quite heavily used for cataloging, it is rather surprising that the ISBN search is quite low. Perhaps librarians search for a similar record and then make modifications.

The very low number of subject searches may be due to the fact that the results obtained will not be meaningful without some form of consolidation of results. Consolidation would require downloading of results from all libraries, and especially in the case of subject searches it could be quite a large task. Another problem is that a unified system of subject cataloging is not yet fully implemented in Poland. An interesting example of subject searches in medical libraries is described later.

### Search target selection

Analysis of how users act shows that about half of all distributed searches are performed by selecting all libraries on the list. On the one hand, library directors are very much in favour of KaRo and support it in every possible way; on the other hand, they are concerned with the load it may generate on their systems. One of their requests is to eliminate the possibility of selecting all libraries with one mouse click. When a user profile that permits selections to be remembered is put in place, this automatic selection will be eliminated.

### User understanding of the interface

Even though much care was taken to make the interface as self-explanatory as possible and help pages are available for every user screen, there are signals that users have problems understanding that timeouts may be due to

too low timeout limits that they can control. Similarly, the low popularity of using KaRo in the single-library mode may come from the fact that only a few users have read the documentation or have experimented with clicking the link representing an individual library.

### KaRo as a back-end system

An interesting use of KaRo was made by Piotr Krzyżaniak, who has set up a WWW interface to Medical Subject Headings (MeSH®). In this system, when users locate a subject heading they are interested in, they can start a distributed search of medical libraries (made by a behind-the-scenes call to KaRo).

## 7 KaRo and NUKat

As we have explained before, KaRo complements the Polish Union Catalog NUKat. It is expected that the catalogs of the main Polish libraries will be loaded one after another into NUKat with duplicate elimination. At this stage, it would make no sense to search these libraries in distributed fashion, when much better quality results can be obtained by searching NUKat. A more difficult situation will arise if only part of the catalog is loaded. Then, in order to get full results, the local library catalog will have to be searched as well, and we will get duplicate hits for those items that are loaded into NUKat. Distributed searches should still be used when the scope is limited to a certain city or library type. Currently, with a standard distributed search, a user is presented with results in a form of a list of libraries along with a number of hits for each of them. When NUKat is also searched, it appears as another library, but the meaning of results is different, as the exact location of the book can only be known after reading the NUKat record. Unfortunately, duplicated information may be received this way, since some books reported in the NUKat search may also appear in results obtained from a direct library search. The only way to eliminate this possible duplication would be to collect all records found in NUKat and analyze them. This would put additional burdens on both KaRo and NUKat servers and would probably be impractical.

It is quite possible that when the NUKat database becomes quite large, most users will access it directly, and the interest in KaRo will disappear. Such a situation will certainly arise with searches made for cataloging purposes, which is quite natural, since the main goal of NUKat is to speed up cataloging and improve its quality. Searching for rare books will probably be quite useful for much longer. At this moment there seem to be numerous reasons to keep KaRo running and develop it further.

## 8 Future Work

### New features already implemented

We have already mentioned some obvious problems with the current implementation, and indicated that some have been already fixed in the new version.

From the KaRo home page, one can access the experimental version currently under development. There are some major differences between the current stable version and the one under development. The most important is the change of the underlying Z39.50 tools, from ZetaPerl to an in-house module based on yaz.<sup>4</sup> Yaz is under constant development, which guarantees that new features can be added to KaRo in the future. In addition, yaz is distributed in the source format and can be modified, for instance to handle servers that do not adhere to the Z39.50 standard in every detail. This change from ZetaPerl to yaz is, fortunately, quite transparent to the user.

One other 'invisible' change is the ability to search and present data in a single network operation, which improves performance. There are also three visible changes:

1. It is possible to save any selection of libraries, so that when one accesses the system again, the selection checkboxes next to the libraries are automatically checked. The option to select all libraries with one click has been switched off, as requested by system librarians. These two

---

<sup>4</sup> Website of Indexdata providing the free yaz toolkit: <http://www.indexdata.dk>.

changes together should significantly lower a number of unnecessarily wide distributed searches.

2. It is possible to select the preferred view type to be either standard bibliographic, bibliographic with holdings or MARC. In addition, whenever a search returns only one hit, the preferred full view is used instead of the brief one. This should be a big help for librarians searching with an ISBN number for a single record. In such a case, setting the MARC view as the preferred one will allow the librarian to get this MARC display directly after clicking on one of the libraries visible on the distributed search result list. Of course, one can always redisplay a record in another view.
3. In the list of distributed search results, a small icon next to the record displays the individual result in another window. This allows the user to keep the list of results in one window and easily change libraries to view various possibilities. This feature is in a very experimental phase and currently may introduce some disruption to the system.

## Plans for KaRo Version 2

The most important new feature of KaRo V. 2 will be the introduction of individual user profiles. Within a profile, a user will be able to save

1. libraries to be visible on the KaRo list
2. libraries to be initially selected
3. preferred settings of timeouts, number of records per page, bibliographic view
4. preferred search fields.

There will be an option for copying profiles to help establish a common core of profiles for all users of some group. There is no plan to store user identities (names) in the profile. Anyone will be able to create an individual entry. The use of a password will be necessary only when changing the profile.

KaRo V. 2 will have a translator of binary MARC to local character encoding. The setting will also be a part of the user profile.

One important internal change will be added: support for storing the Z39.50 configuration in an LDAP directory. The new configuration will have more load limiting parameters, and KaRo will control the overall load on local systems by counting the number of open sessions, searches in progress, etc. For each library or multiple library server, it will be possible to set load limits which free local systems from unwanted searches. This solution may, in some cases, impair KaRo performance, but this is certainly a better option than forcing a library to withdraw from KaRo altogether.

## Part 4

### Hungarian Union Catalogs



## **Chapter 15**

# **The Hungarian Shared Cataloging Project: MOKKA**

Géza Bakonyi

It was only this year that the Hungarian shared cataloging project reached a state, after five years of difficult birth, in which libraries and users could begin to take advantage its services. The main database includes the records of the OPACs of the 15 largest Hungarian libraries: some 1.8 million records net of duplicate records in the database. The database uses authority control on the names, and the records contain the location codes of the member libraries. Through the links related to these codes, we can access the local databases (e.g. for holdings information). The database is updated regularly as material is exported and filled by the member libraries.

A number of special problems proved to be obstacles in the execution of the project.

The first problem was the lack of suitable institutional backing. At the time of the establishment of the project, there was no institution that could provide a financial and professional backing for it; to wit, the National Library had its own problems to cope with, since its own library software was inadequate and it did not have a suitable technology and network infrastructure. Therefore the founding libraries were forced to establish an association for the management of the Hungarian shared cataloging project. Unfortunately, this was not a satisfactory solution either, because it was unable to support the project financially and could not assure professional backing for the project either. As a consequence, at the beginning of 2002, the project was removed from the auspices of the association to the

National Library, and its professional management was replaced as well.<sup>1</sup> The association continues to provide for representation of the interests of the member libraries, and takes care of the operation and development of the project.

The choice of the proper library software was another problem. The vendor selected by tender experienced a crisis and was not able to live up to expectations. As a result, the association was forced to turn to the runner-up. Moreover, the original selection had the further disadvantage that the vendor did not have an agency in Hungary. The new vendor was the Hungarian firm Dataware, and its library software Corvina, originally designed in the USA, had been developed according to the specifications of Hungarian libraries. This library software is used by the largest Hungarian university and public libraries. On the other hand, the vendor was already experienced in building shared cataloging systems, since it had created a cumulative central catalog containing more than 2 million records (prior to deduplication).

It was also problematic that the project did not have a server of its own, which caused difficulties during the development period. However, this year, the National Library concluded an agreement with the Office of the National Information Infrastructure Development Program, which placed one of its servers at the project's disposal. As a result, the project was assured of data storage and sufficient memory capacity, was able to run the software and could cover the payment of the fees for hardware maintenance.

However, the most significant problem was the member libraries' lack of experience in shared cataloging. There was the issue of the quality of the records of the library catalogs, because the main catalog could not solely comprise the customer's own materials. The majority of the member libraries did not have experience in shared cataloging, there were significant differences among local cataloging rules and practices, and the members were at different stages of information technology development. Shared cataloging was made more difficult by the fact that certain libraries used USMARC as a cataloging format, while others used the national

---

<sup>1</sup> At the present time, the author is charged with directing the project.

Hungarian MARC format HUNMARC, and some did not use any MARC formats at all. Moreover, only few of them were able to export their records in any MARC format.

At the time of the transfer of the Hungarian shared cataloging project, under new management, to the National Library this year, there was already a test database, but it failed to meet numerous requirements; for example, problems were encountered when searching the database, displaying hits, etc. During the past few years, the professional staff of the association concentrated on the documentation of the project (Hungarian MARC application rules, rules for the usage of the Hungarian shared catalog, rules of communication with the central database, cataloging codes, etc.). A very good set of materials was created, but a functioning model that would enable the vendor to prepare a fully satisfactory system was still missing. We therefore turned our attention this year to the preparation of such a model.

The bibliographic and export formats used by the libraries caused us the most anxiety. Records can be uploaded to the main catalog in two formats, USMARC or HUNMARC. As mentioned earlier, some of the libraries do not work with any MARC formats; hence, the export of their records in MARC format was not possible at all, or only with numerous syntactical errors. Another problem is that the libraries that use some kind of MARC format do not have identical experience, because they typically use different versions of different MARC formats. Naturally, this caused a special problem in the case of the linked records (e.g. in case of multi-volume items). This problem is handled by us in two ways. On the one hand, before uploading records we use some software that checks the MARC format run. While it is running, a log file is created and an analysis of the error messages in this file permits the creation of filters for modifying the output of the uploads (at least in the phase of initializing the main database). On the other hand, several conversion programs have been prepared, and with these we can convert the bibliographic records of different MARC formats back and forth. Naturally, we had fewer problems with the conversion from HUNMARC to USMARC. In this case, we had only to face the inconsistency arising from the different versions of the MARC formats: first of all, the contradiction in cataloging multi-volume items, identifying local data, filling in the notes fields, etc. The reverse

conversion was more difficult for us, since the Hungarian MARC format is more segmented, does not contain punctuation marks (unlike the USMARC format), the function of the indicators is expressed by subfields, etc. The conversion software plays a significant role in the system, since the default format of Hungarian shared cataloging for downloading, uploading and displaying records is the Hungarian MARC.

The member libraries of the shared cataloging project accepted a cataloging rule that specifies a minimum, obligatory level of cataloging. On the basis of these rules, a filter program was prepared that exercises syntactical control of the records during uploading. Because of the rather liberal interpretation of the cataloging rules by the member libraries (in fact, all of them used their home-made rules), we had to decrease syntactical control when initializing the main database. For instance, not all the libraries use the obligatory fields and subfields (e.g. edition, imprint fields). Moreover, the use of the codes in the different positions of the record heading was also ambiguous. Certain libraries do not indicate whether they do or do not follow the cataloging and punctuation directions of the International Standard Bibliographic Description (ISBD), forget to mark correctly the level of the bibliographic description of the record, and do not mark the place or language of the publication. Unfortunately, we cannot increase syntactical control, since the member libraries become capable only slowly of applying the minimum rules of cataloging. In the case of certain libraries, it is their own library systems that do not support the standard MARC format.

Hungarian libraries began to use computers and networks at the end of the 1980s. It was an exciting and heroic age, but unfortunately it did not pass without leaving some bad legacies. Various character coding tables were used in this period, and if I recall correctly, we used five of them in the member libraries of the MOKKA Project. Regrettably, three different Hungarian character coding tables are used in the libraries even today. Therefore, before checking the MARC format and imposing syntactical control, character conversion must also take place, which creates new sources of errors both in uploading and downloading. Obviously, it is not possible to force the libraries to use the ISO 8859-2 coding table, and the transition to UNICODE also has its difficulties, not to mention that the

vendors of integrated library systems do not seem to be willing to make similar changes in their software.

Ordinary users can access the main database through a WWW interface <http://www.mokka.hu>. They can search and download a limited number of records. I shall omit the details, but we have tried to provide search and display options that are responsive to users' desires. However, we need to emphasize two features. First, it is not only possible to indicate the names of the local libraries in the displayed record; users can also switch to the local OPACs and gain information on local holdings (that is, how many items there are, whether they can be borrowed or where they are located, and so on.) With certain libraries, this linking process emulates the way in which a URL syntax question is sent through the WWW interface of the local database. In other cases, the project supported a local solution or development.

Another piquancy is that the three largest Hungarian libraries have recently started a new project. They have tried to harmonize their subject headings. The thesaurus records of the National Library and the subject heading records of two university libraries were uploaded to a common central database, through which the local catalogs were searchable. In addition, we linked the main database of the Hungarian shared catalog to this system (<http://www.matriksz.hu>).

As everyone knows, a shared cataloging project cannot be declared finished at any one particular moment. Consequently, the Hungarian shared cataloging program is a process, and only its first phase was terminated in the summer of 2002. After this first phase, we can summarize the most important lessons and must specify future tasks. These lessons are not only important from the point of view of the future of the program, but can also determine the obligations of the member libraries.

We are fully aware of the fact that the member libraries face countless problems in their daily routine, and the tasks of shared cataloging will demand that we find solutions to further problems. Nevertheless, we believe that launching and operating the Hungarian shared catalog contain a lesson of vital importance beyond the practical aspects: namely, that the attainment of a high professional level is essential for the development of the Hungarian library services.

## 1 Problems with the Project

Having completed the first phase, it is clear that the group of member libraries must be expanded, but the question is to how many libraries. According to the original plan, we counted on the libraries that participate in the Hungarian document delivery system, i.e. 57 libraries. It is highly probable that only a few of these should be involved in uploading records, since in the rest of the libraries the only new items we have to deal with are items from local historical collections. (Unfortunately, because of the nature of the legal deposit system, not all these items can be found in the catalog of the national library.)

In any event, we must consider the need to realize an effective document delivery system. This implies that location information data of the above 57 libraries must be uploaded. Then queries sent to the union catalog could especially support inter-library loans. Thus, a second step is the realization of the electronic inter-library loan system. However, the shortage of funds makes it difficult to predict when we will achieve this.

At present, the records of the union catalog can be accessed and downloaded (25 records per session) by everyone. The precise rules for downloading and for settling accounts among the members of the project remain tasks for the near future.

## 2 Problems with the Construction of the Union Catalog

As mentioned above, the main format of the union catalog is the Hungarian MARC format, but many libraries use a version of the USMARC. At present, the archival format for the main database does not perfectly handle the formats of the exported records of the member libraries. It is one of our tasks for the current year to find a solution to this problem by cooperating with the vendor.

The treatment of the authority records is not perfect either, because of the great variety of authority files associated with local catalogs. The manual correction of the authority records is conceivable, but it is time-consuming and expensive work. We are still working on how to solve this problem.

According to the original plan, we would like to use, in addition to the union catalog, other authority files in the background (similar to the name files of the Library of Congress).

### 3 Problems of Management and Operation

As mentioned earlier, the member libraries do not stick to the obligatory minimum of the bibliographic description. Accordingly, we plan to set up a permanent committee to clarify this situation.

After analysis of the log files created in the process of uploading the records, we make suggestions to the member libraries about how they can correct errors in bibliographic descriptions and the export of records.

We must refine the process of routine uploading by member libraries.

Switching from the Web interface of the union catalog to local databases in order to get the status information on items is still a problem, as is the reverse process. We must develop a software solution for this.

To summarize, the problems of the Hungarian shared cataloging project arise from the great variety of the Hungarian library system: 15 libraries with different cataloging rules, five different integrated library systems, three different archiving formats, two different MARC formats, etc. But our tasks are clear, and the appropriate steps will be taken by the end of 2002.



# Chapter 16

## Subject Cataloging in a Cooperative Cataloging Environment

### A Case Study

Klára Koltay

The present study discusses particular issues in the area of subject access in a cooperative cataloging environment, and uses as examples three cooperative databases in Hungary: the bibliographic databases of the Hungarian National Shared Catalog (MOKKA) (<http://www.mokka.hu>),<sup>1</sup> the location database of the National Document Delivery System (ODR) (<http://odr.lib.klte.hu>), based on the cooperative cataloging program of the Corvina libraries (VOCAL),<sup>2</sup> and the Matriksz database (<http://www.matriksz.hu>), which consists of three subject heading systems used in Hungary and Universal Decimal Classification number records.

---

<sup>1</sup> Géza Bakonyi, “Mi a MOKKA?” (What is Mokka?), *Könyvtári Levelezőlap*, 10 (1998): 3–5.

<sup>2</sup> Klára Koltay, “VOCAL—a model for union catalog,” *Research Libraries: Cooperation in Automation*, eds. Jadwiga Wozniak and Robert C. Miller (Cracow, November 16–19, 1998: 151–156); Géza Bakonyi, “VOCAL—a Corvina könyvtárak osztott katalogizálási rendszere” (VOCAL—the distributed cataloging system of the Corvina libraries), *Könyvtári Figyelő*, 45 (1999), [http://www.oszk.hu/kiadvany/kf/1999/2/bakonyi\\_h.html](http://www.oszk.hu/kiadvany/kf/1999/2/bakonyi_h.html); László Balázs and Klára Koltay, “Lelőhelyszolgáltatás osztott katalogizálási bázison” (Location database based on cooperative cataloging), paper presented at Networkshop, 1999, Nyiregyháza, <http://www.iif.hu/rendezvenyek/networkshop/99/cdrom/docs.htm>.

## 1 Problems of Subject Access in Cooperative Cataloging Databases

When cooperative cataloging systems start to operate, one of the possible complications is the handling of the various thesauri and subject heading systems used in the cooperating libraries. The complications become real problems to be solved if a cooperative cataloging system not only aims at providing a common pool of bibliographic records for copy cataloging, but also aspires to be open for public use as the common catalog of several libraries or a location database for inter-library loan and document delivery systems.

The problem is especially complicated in Hungary, since for decades the preferred subject access tool of most Hungarian libraries has been the Universal Decimal Classification system, and a comprehensive Hungarian vocabulary has not been developed. The present situation is that nearly all libraries use UDC strings in their catalog records and databases, while only a portion of them add natural language subject terms. Those that use subject headings either employ some in-house system of various levels of vocabulary control or a thesaurus of a limited subject area (Table 1).

**Table 1. Usage of Classification Systems**

	Number of participating local catalogs (July, 2002)	Number of libraries using				
		UDC*	Other classification systems	Local subject terms	LCSH** (English and/or Hungarian)	MeSH***
MOKKA	16	11	1	10	2	2
ODR	11	10	-	7	2	1

\* Universal Decimal Classification

\*\* Library of Congress Subject Headings

\*\*\* Medical Subject Headings

A plausible way of handling the various types of subject information provided by local catalogs is to enter them in the relevant union database records without paying too much attention to the existence of the various heading systems, and to use the natural language terms as one set of uncontrolled subject keywords and the UDC strings as the basis for classification number searches.

Both MOKKA and ODR, the bibliographic databases of our study, have chosen the above approach: the MARC bibliographic records contain all relevant locations, subject terms and UDC strings added to it in various libraries (Table 2). The subject heading fields may contain coded references to their system of origin. One slight philosophical difference is that the MOKKA database, being consistent with the compromise already made, does not undertake to store the reference systems of the various subject schemes and uses only skeleton authority records in the case of subject fields. On the other hand, the VOCAL database uses the original, full-subject authority records of the member libraries, in the hope that the reference information in them will enhance the accuracy of subject browse searches even in a mixed subject environment.

**Table 2. A Typical ODR Record with UDC Strings, Subject Terms and Location Information**<sup>3</sup>

```
000 01156nam 2200241 i 4504
001 bibFSZ724579
005 20020723154313.0
008 s2002 hu 0 hun d
020 ♯a963-9376-46-9 (kötött)
040 ♯aHuBpFSZEK♯dSz1/91♯dHudEKLEK
080 0 ♯a931(089.3)
080 0 ♯a930.85(3)
080 ♯a930.85(3)(089.3)♯a931(089.3)
```

---

<sup>3</sup> UDC strings are 080 fields, subject terms are 650/651/695 fields, and location information is 949 fields. The various subject schemes present are defined by the indicators of the subject fields.

100 1 \$aNémeth György\$d(1956-)\$c(történész)  
 245 10 \$aKarthágó és a só :\$baz ókortörténet babonái  
 /\$cNémeth György  
 260 \$aBudapest :\$bKorona,\$c2002  
 300 \$a215 p. :\$bill. ;\$c24 cm  
 504 \$aBibliogr.: p. 206-215. és a jegyzetekben p. 201-205.  
 520 \$a"Mindenki úgy tudja ... , hogy Karthágót a rómaiak  
     porig rombolták, majd a helyét sóval behintették, hogy  
     Drákón hírhedetten kegyetlen törvényeket hozott ... és  
     hogy Attilát hármas koporsóban temették el a Tisza  
     medrében. És mindenki rosszul tudja!"  
**650 4 \$aTörténet, ókori\$xtévedések, koholmányok stb.**  
**650 0 \$aHistory, Ancient\$xErrors, inventions, etc.**  
**650 7 \$aMűvelődéstörténet\$y ókor\$xkuriózumok**  
**650 7 \$aKuriózumok\$xókori**  
 949 \$1D1/CA  
 949 \$1D1  
 949 \$1S21  
 949 \$1S21/91  
 949 \$1S24/K2  
 949 \$1S24/F2  
 949 \$1D1/OR  
 949 \$1B10/X  
 949 \$1Sze1

Despite the fact that the information is present in the records (coded with the help of indicators or source subfields in the subject fields), the retrieval mechanism of databases disregards the origin of subject terms in both keyword and browse subject searches. The keyword search strategy we employ must be the one we would use in an uncontrolled subject keyword environment in which a concept can be described in different ways, where various synonyms and endings can appear and a set of records does not contain subject terms at all.

## 2 Search 1

The results of a test search for the concept of deviant behavior, given in Table 3 below, proves the dominance of UDC strings as a subject access tool in both databases, which are probably not very often used by online searchers. The possible subject terms, on the other hand, can vary considerably within a database.

**Table 3. Subject and UDC Search Results in Some Union and Individual Library Catalogs**

Search term	Index	Number of records in			
		MOKKA	ODR	DEENK*	SZTEEK**
316624%	UDC	132	256	26	-
34395%	UDC	79	123	8	-
<i>deviancia</i>	Subject	98	87	4	4
	Subject or title	126	137	12	-
<i>Deviáns viselkedés</i>	Subject	15	76	11	6
	Subject or title	20	85	12	-
<i>Aszociális viselkedés</i>	Subject	0	1	0	0
	Subject or title	0	1	0	-
<i>Beilleszkedés i zavarok</i>	Subject	21	23	3	32
	Subject or title	21	23	14	-
<i>Bűnözés</i>	Subject	245	467	53	83
	Subject or title	398	673	95	-

Search term	Index	MOKKA	ODR	DEENK*	SZTEEK**
<i>Bűnözői viselkedés</i>	Subject	2	4	2	0
	Subject or title	2	4	4	-
<i>Antiszociális viselkedés</i>	Subject	0	3	1	-
	Subject or title	0	4	2	-
UDC 316624% or subject <i>deviancia</i> or title <i>deviancia</i>		245	332	33	-

\* Debrecen University Library

\*\* Szeged University Library

A few plausible ones used in the test searches, the selection of which depends entirely on the searchers' creativity and foreknowledge of the subject schemes of the databases, came up with increasingly new results. It is only if we try to construct a more composite search, which is again highly unlikely to happen with lay users, that a fuller result emerges (Table 3.).

Though the subject access provided in this way may seem sufficient when our aim is only to find a few titles on a certain subject, it can seldom give a comprehensive result and cannot help users with all the guidelines that are built in the reference system of thesauri and controlled subject heading systems.

### 3 Subject Databases

However, the information present in the databases and the subject authority records allows us to complement the subject access method described above with a more sophisticated one, which preserves the integrity of each vocabulary for those who want to make use of their reference system, and creates parallelisms among the terms used for the same concept in the various vocabularies for the users who want to search across them.

After some technical experiments conducted separately at the National Széchényi Library, with its thesaurus database, and in the Szeged and Debrecen University Libraries, with a subject heading database attached to the VOCAL cooperative cataloging system,<sup>4</sup> a consortium of the National Széchényi Library, Szeged University Library, Debrecen University Library, the Kaposvár County Library and the Library Institute was founded in 2001 to address the above problem and to offer a service which, even in the absence of a single, generally-used subject heading system, can give guidance both for subject catalogers and for searchers nationally.

The first phase of the project, called Matriksz, was completed in March, 2002 and concentrated on working out the technical framework using the thesaurus of the National Széchényi Library, the Szeged subject headings, the Library of Congress Subject Headings, translated into Hungarian at the Debrecen University Library, and the UDC index tables, translated by the Library Institute. The Matriksz database is currently available for real-life tests.

In its present state, the Matriksz service consists of a fully searchable database of subject headings and classification numbers stored in MARC format records, displaying and allowing navigation according to the reference structure of headings, and maintaining parallelisms among the heading systems and between the headings and UDC strings. In Figure 1, the left-hand panel shows the integrated list of results from the four resource subject schemes. The sources of terms and classification numbers are indicated in brackets with index numbers referring to the number of existing term and subdivision combinations. In the case of “*Deviáns magatartás és szubkultúra*” (deviant behavior and subculture), the two combinations of terms and geographical subdivisions are displayed. The right-hand panel shows the selected item of the result list in full display. All references and equivalent terms and UDC numbers are points of further navigation. They are links, activated by clicking to make the system perform another search with the selected term and display its environment.

---

<sup>4</sup> Klára Koltay, “Az ODR adatbázis új szolgáltatásai,” (New services by the ODR database), *Tudományos és Műszaki Tájékoztatás*, 48/8(2001): 315–321. The database (with the nickname termdb) is accessible at [http://vocal.lib.klte.hu/corvina/opac/term\\_search](http://vocal.lib.klte.hu/corvina/opac/term_search).

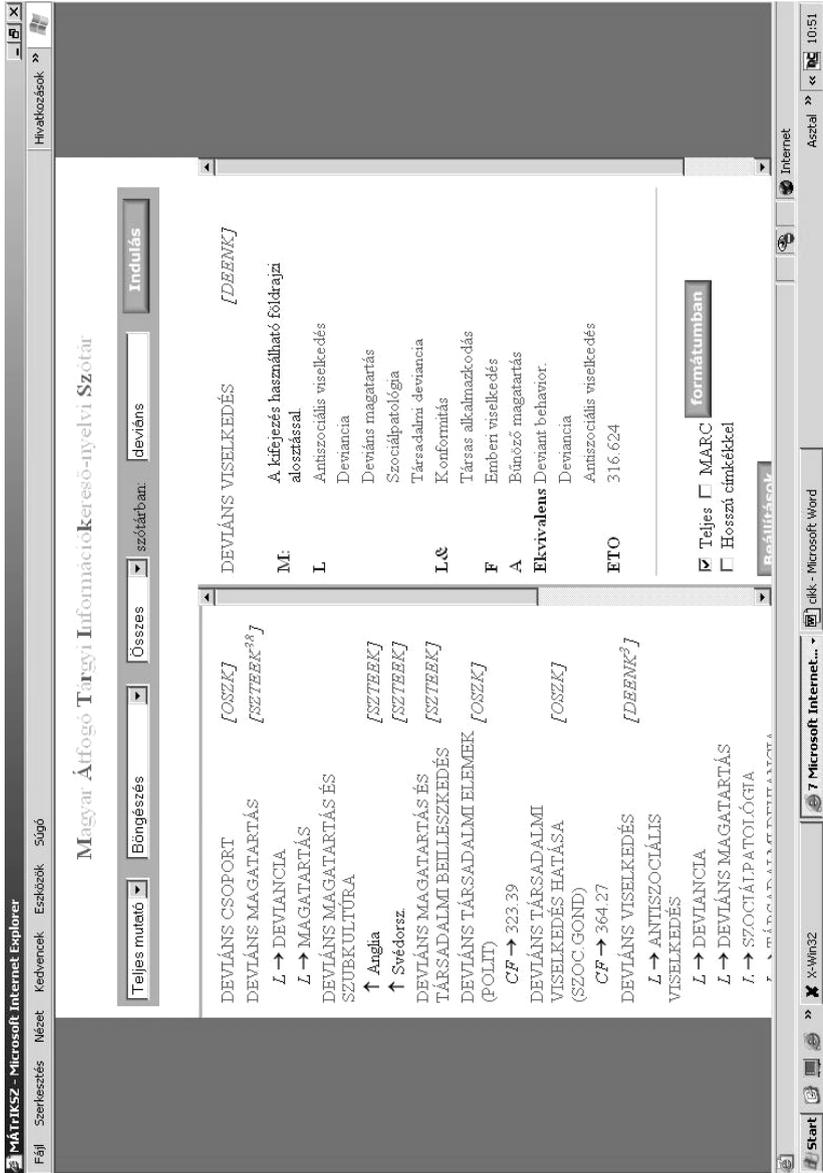


Figure 1. The Result Screen of the Browse Search “deviáns”

[MÁTRIKSZ - Microsoft Internet Explorer](#)  
 Fájli Szerkesztés Nézet Kedvencek Emlékek Súgó

[MÁTRIKSZ - Microsoft Internet Explorer](#)  
 Fájli Szerkesztés Nézet Kedvencek Emlékek Súgó

Magyar Átfogó Tárgyi Információkereső-nyelvi Szótár

Teljes mutató Böngészés Összes szótárban: antiszociális Indulás

antiszociális személyiségzav  
 antiszociális viselkedés  
 deviancia  
 deviáns csoport  
 bűnöző megatartás  
 34395

DEFNK  OSZK  SZTEEK  VOCAL  
 Cím  FTO  Tárgyszó  
 C és G vagy **Keresés**

**Térítés**  
**Elég**

343.95  
*M:* Bűnügyi lélektan. Kriminápszichológia. Igazságügyi lélektan  
 CA → BŰNÜGYI LÉLEKTAN  
 CA → IGAZSÁGÜGYI LÉLEKTAN  
 CA → KRIMINÁLPszichológia

343.96  
*M:* Bűnügyi patológia. Kriminápatológia. Elmebetegség és bűnözés. Felelősség és beszámítóképesség  
 CA → BESZÁMÍTKÉPESSÉG (KRIMINOLÓGIA)  
 CA → BŰNÜGYI PATOLÓGIA (KRIMINOLÓGIA)  
 CA → FELELŐSSÉG (KRIMINOLÓGIA)  
 CA → KRIMINÁLPATOLÓGIA

343.97  
*M:* Bűnügyi szociológia. Kriminászociológia. A bűnözés mint társadalmi jelenség. A bűnözés okai. A társadalmi környezet befolyása. Összehasonlító bűnügytan  
 CA → BŰNÜGYTAN

[O D]  
 [O SZK]  
 [O SZ]

Figure 2. More Results from Browse Search

Following these links, users navigate through a chain of terms related in one way or another to their concept of interest and can pick the ones they feel are more relevant to their present needs. Thus, the Matriksz database allows users to select the ‘proper subject terms’ before they start their subject query in bibliographic databases.

The second element of the service offers a number of bibliographic databases that can be searched with the selected subject terms one by one or collectively. In Figure 2, the left-hand panel again shows a segment of the result list, while on the right some of the terms and UDC numbers that might be relevant for a bibliographic search in our selected subject area are already collected. One or several of the target bibliographic databases can be selected, as well as the types of searches (title and/or UDC and/or subject) we intend to perform.

**Table 4. a. Debrecen record; b. Szeged record; c. OSZK record; d. UDC record**

```

0759nz 2200265n 4504
001 autKLT00204617
005 20020730132045.0
008 971218nn acnnnbabn un aaa d
040 $aHuDeKLEK
080 $a31b.624
50 4 $aDeviáns viselkedés
450 4 $aAntiszociális viselkedés
450 4 $aDeviancia
450 4 $aDeviáns magatartás
450 4 $aSzociálpatológia
450 4 $aTársadalmi deviancia
550 4 $wh$aBűnöző magatartás
550 4 $aKonformitás
550 4 $aTársas alkalmazkodás
550 4 $wg$aEmberi viselkedés
680 $aA kifejezés használható földrajzi alosztással.
690 $xJE, OSZK
750 0 $aDeviant behavior.

```

750 7 \$aDeviancia#20SZK  
 750 7 \$aAntiszociális viselkedés#20SZK  
 a./

00339nz 2200145n 4504  
 001 autJAT00331084  
 005 20020304082332.0  
 008 970411nn acnnnbabn un aaa d  
 040 \$aJ  
 080 \$a316.624  
 150 7 \$aDeviáns magatartás  
 450 7 \$aDeviancia  
 450 7 \$aMagatartás#xdeviáns  
 690 \$ags#x0SZK  
 750 7 \$aDeviancia#20SZK  
 b./

## DEVIANCIA

000 00733nz 2200277n 4500  
 001 0SZK000000053331  
 005 20020514235905.0  
 008 020514 b an n a a  
 040 \$a0SZK\$bHU  
 150 \$adeviancia  
 154 \$a316.624  
 450 \$wy\$aantiszociális viselkedés  
 450 \$wy\$aadeviáns viselkedés  
 550 \$wg\$abeilleszkedési zavar  
 550 \$wf\$alélektan  
 550 \$wf\$aszociálpatólógia  
 550 \$wp\$aalkoholizmus  
 550 \$wp\$abünözés  
 550 \$wp\$akábítószer

```

550   $wp$aprostitúció
550   $wp$aszkinhed
550   $wr$acsoportszociológia
550   $wm$adeviáns csoport
550   $wm$aerkölcs
550   $wm$afiaatalkorú
550   $wm$aszocializáció
c./

000 00230nz 2200097n 4500
001 0SZK0027508
005 20021009235905.0
008 021009 b an a a d
040   $a0SZK$bhu
080   $a31b.624
150   $aszociálpatológia (szoc)
d./

```

The UDC strings are integrated into the system in two ways. The Debrecen subject heading records contain parallel UDC numbers that can act as points of further navigation in the database (Table 4a). At the same time, the index of the UDC's medium edition is entered in the form of classification number records containing the numbers and their definition (Table 4d).

Being a MARC database, it can be indexed according to various rules in a flexible way. The present Matriksz database and its predecessor, the VOCAL subject database nicknamed termdb, represent two different approaches.

Termdb uses one big keyword index containing the headings (field 150) and all the references (fields 450, 550, 750) and notes (field 680) in its default search. It gives maximum guidance to users who do not really know which terms are the accepted headings, though it sometimes returns too many hits to be really helpful.

The Matriksz database indexes only the heading field (150), and the references appearing in the displayed record serve only as points of further navigation. This way, the search results always seem more manageable, although if our search term is only a 'see' reference we might miss important results to start with.

#### 4 Search 2

In Search 2, the task is to use the subject databases for collecting as many relevant terms as possible to describe the subject area of 'antisocial, deviant behavior' of Search 1. In order to find out if there are any real differences due to the different indexing rules, the searches are carried out in both termdb and Matriksz. (Note that the Matriksz database is richer in context, and termdb does not contain the OSZK thesaurus or the OSZK UDC classification records.) Only the harvested terms are listed here.

**Table 5. Terms Collected in termdb**<sup>5</sup>

316.624	
349.95	
Antinomikus személyisé	(antinomian personality)
Antisocial personality disorders	
Antiszociális személyiségzavarok -- nevelés	
Antiszociális személyiségzavarok	(antisocial personality disorders)
Antiszociális viselkedés	(antisocial behaviour)
<b>Bűnözés</b>	<b>(crime)</b>
<b>Bűnöző magatartás</b>	<b>(criminal behaviour)</b>
<b>Bűnözői viselkedés</b>	<b>(criminal behaviour)</b>
Deviancia	(deviancy)

---

<sup>5</sup> Terms collected through three searches and by consulting the full records displays.

Deviáns magatartás	(deviant behaviour)
Deviáns viselkedés	(deviant behaviour)
<b>Emberi viselkedés</b>	<b>(human behaviour)</b>
<b>Ifjúságszociológia -- deviáns magatartás</b>	<b>(sociology of young people - deviant behaviour)</b>
<b>Konformitás</b>	<b>(conformity)</b>
<b>Önmegsemmisítő magatartás</b>	<b>(self-destructive behaviour)</b>
Pszichopatologikus személyiség	(psychopathic personality)
Személyiségzavarok	(personality disorders)
Szociálpatológia	(social pathology)
Szociopatologikus személyiség	(sociopathic personality)
Társadalmi deviancia	(social deviancy)
Társas alkalmazkodás	(social adjustment)
Társas készségek	(social skills)

**Table 6. Results of the Same Search in the Matriksz Database**

316.624

**323.39**

364.27

antinomikus személyiség

antisocial personality disorder

antiszociális személyiségzavarok

antiszociális viselkedés

aszociális viselkedés (szoc)

deviancia

**deviáns csoport**

deviáns magatartás

**deviáns magatartás és szubkultúra**

**deviáns magatartás és társadalmi beilleszkedés**

**deviáns társadalmi elemek (polit)**

**deviáns társadalmi viselkedés hatása (szoc.gond)**

deviáns viselkedés

deviáns viselkedés

**marginális viselkedés (szoc)**  
**marginális, aszociális viselkedés**  
pszichopatologikus személyiség  
személyiségzavarok  
szociálpatólógia  
szociopatologikus személyiség  
társadalmi deviancia

(and some not fully relevant terms, primarily owing to the “UDC index” records).

## 5 Searching the Bibliographic Databases

Whichever of the above indexing and search methods we prefer, it is very useful for subject searchers to get acquainted with the expressions and UDC numbers employed by our databases. Finding a relevant subject term might in itself be very helpful for subject catalogers, but for patrons who are interested in getting information on books and their locations, that is just a first step. They would want to use the selected subject term to search the bibliographic database of their choice.

The Matriksz project has put great emphasis on providing this service at the present time for the OPACs of the member libraries and for the VOCAL/ODR database. (It will soon be available for the MOKKA database, as well.) One or a combination of databases can be selected as a target for bibliographic searches. The default provided at present is the combination of the VOCAL/ODR and OSZK catalogs, which presents the widest possible range for searches at this time: the 11 full databases and additional location information from 45 ODR libraries, and the catalog of the National Széchényi Library.

It is quite straightforward for the user to switch from searches performed in the Matriksz database to bibliographic searches in the chosen remote online catalogs. As is shown in Figure 2, during one or several Matriksz database searches users collect their terms of interest: these can either be real language expressions or UDC strings (truncated as much as relevant), which, owing again to the parallelisms established in the subject and call number records, are much easier to interpret than in an average UDC

search. Users then select the target bibliographic database and indicate which indexes of the target database they wish to search.

The indexes that are offered on the Matriksz interface for searching the bibliographic database are not only subject and UDC indexes. In an environment in which there exist major catalogs without subject terms, it is also important to use title fields and indexes when performing subject searches.

## 6 Search 3

The expressions and UDC strings collected in Search 2 are now used in bibliographic searches. The subject terms relevant to the topic and revealed through navigation in the subject databases in Search 2 are now collected for one complex search in various target bibliographic databases. The search terms used are all of the following: *deviancia, deviáns viselkedés, antiszociális, viselkedés, deviáns magatartás, beilleszkedési zavar, bűnözés, bűnözői viselkedés* (deviance, deviant behavior, antisocial, behavior, deviant conduct, difficulty in adaptation, criminality, criminal behavior).

**Table 7. Search Results from Search 3**

	Number of records in the catalogs of				
	DEENK	OSZK	SZTEEK	VOCAL	VOCAL+OSZK
<b>Subject search</b>	70	23	169	712	735
<b>Subject or UDC</b>	93	26	216	972	998
<b>Subject or UDC or title</b>	140	33	245	1000	1033

The search results, when compared to those in Table 3, permit the conclusion that the one composite search formulated with the aid of the Matriksz database resulted in more records than the several searches performed in Search 1. Looking at the data from the Szeged and Debrecen catalogs suggests that Matriksz can even serve as an enhanced subject

search tool for library catalogs, compared to the standard OPAC subject search services.

The results of the bibliographic searches appear alphabetically in the Matriksz search result screen, in a single list with an indication of the resource database. Records can be viewed one by one in longer formats. The MARC format is especially important here, since it is the only source of location information at present. As the databases themselves may contain sundry location information and other services, it will be important in the future to enable Matriksz to lead us not only to its own result screens, but back to the interfaces of the searched bibliographic databases (Figure 3.). The left panel displays a segment of the united search result, with the name of the source database in brackets. The right panel shows one of the VOCAL records displayed in MARC with the holdings library codes.

## 7 Database Maintenance

The four elements of the Matriksz database differ considerably in their maintenance as well. The records of the UDC index list and the OSZK thesaurus can be considered relatively complete. The occasional updates and additions are primarily processed with an in-house thesaurus management software called Relex,<sup>1</sup> and the results converted to MARC format are fed into the Matriksz database periodically. The Debrecen<sup>2</sup> and Szeged subject lists are developed during daily cataloging work and the Matriksz database itself, together with the libraries' ILS cataloging modules, is relied on heavily when subject catalogers have to decide which terms, and in what form, can be added to the existing headings without disturbing the coherence of the system. The new headings and updates to

---

<sup>1</sup> For the OSZK thesaurus and its maintenance, see Rudolf Ungváry, "Az OSZK tezauszusa és a KÖZTAURUSZ," *Könyvtári Figyelő*, 47 (2001).

See [http://www.oszk.hu/kiadvany/kf/2001/1/ungvary\\_1.html](http://www.oszk.hu/kiadvany/kf/2001/1/ungvary_1.html).

<sup>2</sup> Klára Koltay, "Why and how to translate a subject heading system?" in *Library Automation in Transitional Societies: Lessons from Eastern Europe*, eds. Andrew Lass, and Richard E. Quandt (New York: Oxford University Press, 2000: 267–83).

[M](#) [Á](#) [T](#) [R](#) [I](#) [K](#) [S](#)  
 ETO jelzet | Kulcsszavas keresés | Összes | szótárban: 31662\* | [Közzététel](#)

Ferge Zsuzsa: A jóléti állam pszichológiája	[vocal]	40011016104250.U 9712080c19971997hu	b	00000	hun
Ferge Zsuzsa: A jóléti állam pszichológiája	[vocal]	963 8128 94 8 (fűzött)	1200,- Ft		
Ferge Zsuzsa: Van-e negyedik út?	[vocal]	1	hun, eng	159,942	159,923
Forgacs József: A társas érintkezés pszichológiája	[vocal]	1	Goleman, Daniel		316,628
Forgacs József: A társas érintkezés pszichológiája	[vocal]	10	Emotional intelligence (magyar)		
Forgacs József: A társas érintkezés pszichológiája	[vocal]	10	Emotional intelligence / Daniel Goleman		
Forgacs József: A társas érintkezés pszichológiája	[vocal]	1	Budapest : Háttér, cop. 1997		
Forgacs József: A társas érintkezés pszichológiája	[vocal]	1	453 P. ; 18 cm		
Forgacs József: A társas érintkezés pszichológiája	[vocal]	1	(Lélek kontroll, 1217-0364)		
Forgacs József: A társas érintkezés pszichológiája	[vocal]	1	Bibliogr. a fejezetek végén		
Forgacs József: A társas érintkezés pszichológiája	[vocal]	4	NR, KA, JE		
Forgacs József: A társas érintkezés pszichológiája	[vocal]	4	Erzelmeik és megismerés		
Forgacs József: A társas érintkezés pszichológiája	[vocal]	4	Erzelmeik társadalmi szempontok		
Forgacs József: A társas érintkezés pszichológiája	[vocal]	0	Emotions.		
Forgacs József: A társas érintkezés pszichológiája	[vocal]	0	Emotions and cognition.		
Forgacs József: A társas érintkezés pszichológiája	[vocal]	0	Emotions Social aspects.		
Forgacs József: A társas érintkezés pszichológiája	[vocal]	1	Kiss Zsuzsa (1955-)		
Forgacs József: A társas érintkezés pszichológiája	[vocal]	0	Lélek kontroll		
Forgacs József: A társas érintkezés pszichológiája	[vocal]		D1		
Forgacs József: A társas érintkezés pszichológiája	[vocal]		D9		
Forgacs József: A társas érintkezés pszichológiája	[vocal]		P4		
Forgacs József: A társas érintkezés pszichológiája	[vocal]		D856		
Forgacs József: A társas érintkezés pszichológiája	[vocal]		Sze1		

[Gömb](#) [Start](#) [7 Microsoft Internet...](#) [cikl - Microsoft Word](#) [X-Win32](#) [Internet](#) [Asztal](#) [Ugrás](#)

Figure 3. A Bibliographic Search by UDC String “31662%” in Both the OSZK and VOCAL/ODR Databases

the references of already existing headings are created in the local cataloging system and primarily saved to the local authority controlled catalog databases. However, since termdb and Matriksz provide special search capabilities in comparison with the local catalogs important for subject catalogers, it is important for all updates to enter these databases as well. The default save option of the local catalog modules provides the additional functionality of also sending updates to these common subject databases. The method is very convenient and straightforward for adding new information to the subject databases, but cannot really solve the problem of deleting complete subject heading records. This is also an important point of further development.

## 8 Results and Plans for the Future

In an experimental phase, the Matriksz database has charted a possible way of handling subject access in an environment of multiple indexing languages. Instead of aiming for the construction of a new comprehensive vocabulary, trying to persuade database owners to abandon their previous practices, and giving up on the problem of the few million records already indexed with other subject tools, Matriksz tries to work with what is available: it aims at storing the various subject schemes, creates parallelisms among the most important schemes, offers their records for download, thus promoting their use, and develops a search tool that counteracts the confusion created for library users by the existence of several schemes.

The Matriksz database is ready to incorporate other subject schemes as long as they are in MARC format, and ready to extend its accessible bibliographic database list with Z39.50 compliant catalogs/databases.

Besides widening the scope of databases linked to it, the program has four areas in which it plans to enhance its services.

## 9 Developing Vocabularies

Though the individual schemes have resources for development based on member libraries, creating parallelisms requires extra efforts from the

project. As the first tests show, it can be an effective tool for unification from the searcher's point of view.

Given the requirements of building and unifying vocabularies, the Matriksz database will gather detailed search statistics. Since the search terms with which users attempt to formulate their inquiries are often very different from those used in the controlled vocabularies, the project attempts to log the search terms that have been entered, and to process them in such a manner that popular terms entered by users can be identified and built into the controlled vocabularies either as references or as new headings.

## 10 Changing the Indexing Strategies

It is clear from the test runs on the termdb and Matriksz databases that they are indexed differently, and that an optional indexing structure has to be worked out, which will facilitate the unification of the two databases.

## 11 Enhancing Links to Bibliographic Databases

With the growing number of databases linked to Matriksz, it becomes more and more important to create a direct link from the Matriksz bibliographic search result window and its detailed item information and other functionalities to the local catalogs. If we want to use Matriksz as an additional subject tool attached to bibliographic databases, it must lead us back to the target database completely, just as termdb does. This is especially important in the case of location information and ILL services of the ODR database.

## 12 Working out an English-Hungarian Bilingual System

Parallelisms between Library of Congress subject headings and their Hungarian translations already present in the Debrecen records, and the planned English translation of the OSZK thesaurus provide the raw material for an English-Hungarian bilingual system. A well-structured system of

appropriate search indexes in the subject database and a set of rules for the formulation of searches in the bibliographic databases makes it possible for a foreign user to pick English subject terms in Matriksz, which enables searches of Hungarian bibliographic databases using Hungarian subject headings equivalent to the selected English terms. Or, conversely, Matriksz will enable Hungarian users to search foreign databases with Hungarian subject terms.



# Chapter 17

## Principles of a National Union Catalog: Shared Cataloging in a Small Country

Erik I. Vajda

The main aim of this paper is to outline some specific characteristics and the background ideas for some decisions concerning the establishment of MOKKA<sup>1, 2</sup>, the Hungarian National Shared Cataloging System. We assume that the discussion of some of these ideas and of the resulting decisions may contribute (first and foremost, but not exclusively) in smaller countries to the development or improvement of similar systems in the national environment, i.e. shared cataloging systems with the participation of major libraries and national union catalogs as the product of the shared cataloging.

Some of the more or less system-specific characteristics and considerations leading to these decisions are related to the peculiar features of the Hungarian library environment. However, it might eventually also be useful

---

<sup>1</sup> MOKKA is the acronym for the Hungarian name of the Hungarian National Shared Catalog (*Magyar Országos Közös Katalogus*). See also <http://www.mokka.hu>.

<sup>2</sup> There are many papers available about shared cataloging and union catalogs in Hungary, including those dealing with MOKKA. However, these papers are all in Hungarian, and therefore no references are given, except for a single one about the problems of subject searches in a shared cataloging environment. For a general introduction to MOKKA, reference is made to the website of MOKKA (also in English) in general, and to the page <http://www.mokka.hu/e-bemutat.html> in particular. This page describes the history, aims, functional model, structure and possible future of MOKKA.

for some other smaller countries to get acquainted with these decisions and their background, whereas other decisions are based on considerations that seem to be relevant for most national shared cataloging systems, independent of the size of the country. In the first part of this paper, we try to give a survey of these characteristics, whereas in the second part we analyze in detail the question of physical versus virtual union catalogs; a broad question that has been discussed intensively in Hungary as well.

At the outset, one has to realize that the establishment of a shared cataloging system and of a union catalog starts in an environment of libraries with various traditions, habits, computerized library systems etc. However, it is the common interest of all libraries (whether participating in a national system or not) and of their users to have a tool for retrieval from the stocks of all the libraries, the holdings of which cover the majority of titles available in the country. In the paper, we discuss both the major problems and their possible solution.

## 1 Size of the System

The optimal size of a shared catalog and of the national union catalog system can be defined only on the basis of an analysis of goals to be achieved. These goals are:

- To create a tool for libraries and library users that enables them to determine the libraries in which they can find and borrow, or get a copy of, a given document available somewhere in the libraries of the country, but not available in the library in which this demand originated;
- To simplify the processing (cataloging) of documents by copying/downloading items of existing records; and
- To contribute to the use of common standards and standard-like solutions for cataloging and retrieval.

Statistical investigations reveal that it is not necessary to include the catalog data of all libraries in a country, or even of the majority of libraries, in the shared cataloging system (i.e. in the national union catalog). In Hungary, it was proved by such investigations (based on an existing manual union catalog of documents published abroad) that these goals could be achieved

to a great extent by a shared cataloging system of fewer than 20 libraries. As a matter of fact, the 20 or so libraries in this set are, in any event, the main suppliers of documents in inter-library loan and copying services, without taking into account the great county public libraries.

If, however, the coverage of titles needs to be extended even further, the inclusion of the catalog data of more libraries and/or virtual solutions—i.e. the near-completion of the physical union catalog by adding more data from other physical or virtual shared catalogs—can further improve the coverage.

In Hungary, 17 libraries (now actually only 15, because of the merger of four member libraries into two libraries) hold about 70% of all foreign titles available in Hungary and nearly 100% of Hungarian titles. These libraries are the members of the MOKKA system.

## 2 Coexistence—Not Always Peaceful—of Different Library Systems and Standards within One Shared Cataloging System

It is characteristic, with some exceptions, of most (smaller or larger) countries in Eastern and Central Europe that library automation started with the acquisition and use of different automated library systems. For example, the 15 member libraries of MOKKA even use different automated/integrated library systems. The central system of MOKKA uses one of these systems, CORVINA, a version of which had been further developed for the purposes of MOKKA.

Obviously, the use of a variety of systems by the libraries that supply data to the central database causes a lot of problems. Solutions might be the application of the Z39.50 standard, the up- and downloading of MARC records, or the use of, or conversion to, other common standards. MOKKA decided on a solution based on MARC export and import, since the cataloging modules of the overwhelming majority of the library systems used by the member libraries are MARC-based or are at least able to export and

import MARC records (in the case of MOKKA, either HUNMARC<sup>3</sup> or USMARC). In some cases, MOKKA supported the development of in-house tools to facilitate MARC export and import. As a result of these developments, all library systems are able to upload (see the reasons below) and download either HUNMARC or USMARC bibliographic and authority records.

There are also a lot of other conventions (resulting in the use of many conversion programs), because of the diversity of practices and rules among member libraries. They include the following:

- An USMARC and HUNMARC conversion program was needed to convert the record of the uploading library to the internal format of the system, and conversely, conversion programs were needed for downloading in order to convert from the internal format to the MARC-format used by the downloading library;
- The member libraries use various coded character sets, and therefore a conversion of the input to ANSEL (used as the character set of the central MOKKA database), and a conversion of the output to the character set of the downloading library was needed;
- MOKKA (the central database) uses a standard record-linking technique for volumes, and for the whole document in the case of multi-volume documents. Some member libraries use repeatable fields for the volume data, and therefore conversion programs were needed for uploading and downloading record(s) of multi-volume documents, if the library did not use the standard record-linking techniques.

Experience has shown that the above-mentioned problems can be solved, although not easily. Without these solutions, however, consistency cannot be ensured, because it was and is impossible to force a retrospective change of the systems and standards used by the member libraries.

---

<sup>3</sup> HUNMARC, the Hungarian standard exchange format, is USMARC-based, but—mainly because of specific features of the Hungarian language, such as the form of names of persons, but also for other reasons—it deviates from USMARC. Its newest version also takes into consideration the developments included in MARC21. The MOKKA system allows conversion from and to HUNMARC and USMARC.

### 3 Uploading vs. Cataloging in the Central Database

The 'classic' method of shared cataloging applies the following model. When a new document arrives at the cataloging department of the library,

- A search is executed in the database of the shared (union) catalog;
- If the search result is positive, the relevant record and the corresponding authority record(s) are downloaded to the catalog of the cataloging library and completed by local data;
- The name (code) of the downloading library is marked in the union catalog;
- If the search result is negative, the library executing the search catalogs the item in the central database and downloads the record that has been prepared.

The regular MOKKA procedure deviates from this well known practice. Of course, the process in MOKKA also starts with a search of the central database of the union catalog for the item to be cataloged. If the bibliographic record of the item is available in the central database, the cataloging library downloads the record, edits it by adding the contents of fields of local significance (e.g. subject headings, indices of classifications, notes, uniform titles, if not present in the downloaded record, etc.) and uploads the record 'back' into the central database. The uploaded record will be eliminated by a duplication check mechanism except for the identification data of the record-supplying library and its record identifier, as well as for the contents of some fields/subfields (e.g. subject headings, classification indices, uniform titles, country code, notes etc.). These will be added to the records, if different from the content of the given field present in the existing 'central' record.

If the record of the item to be cataloged is not present in the central database of the system, the member library does not catalog in the central database of the union catalog. As mentioned already, the member libraries use different automated library systems. Therefore a number of special cataloging clients would be needed in the member libraries for cataloging in the database of the union catalog, and all catalogers in the member libraries would have to learn the rules of the cataloging modules of both their home library systems and of that of the central system. To avoid the

additional costs and the additional workload on their catalogers, the member libraries of MOKKA decided that their catalogers should not catalog the new item in the central database of the union catalog, but catalog it 'at home,' and upload their records (of course in MARC format) by placing the item in a file designated for this purpose, the content of which is regularly checked by a program for new records in the file. This program uploads the new records to the central database. The duplication check mechanism provides for the elimination of duplicates (although there is usually no duplication, because the cataloger is obliged to check before cataloging whether the given record is not already available in the central database) and if the uploaded document was a duplicate, then only the record identifier, the identification mark and name of the uploading library (and the new contents of some fields mentioned above) are added to the record existing in the central database.

#### 4 Authority Control

The existence and variety of authority files vary from member library to member library. The central database of MOKKA includes authority files for names of persons and names of corporate bodies (including the names of conferences, other meetings, fairs, etc.). In addition, there are formal authority files for titles and subject headings (but not for standardization of the data, only to facilitate their global change if necessary).

For the 'real' authority files, MOKKA uses the following procedure:

1. Libraries that maintain authority files (a minority of cases) have been asked to upload these files prior to uploading the related bibliographic record;
2. The uploaded authority records are placed in the given authority file of MOKKA, following their duplication check and then linked with the relevant bibliographic record;
3. If the given authority data in the uploaded bibliographic record 'find' their authority record, they are linked with each other;

4. If the authority data of uploaded bibliographic record are not present in the authority files, a so-called ‘skeleton’ authority record is prepared and linked to the given bibliographic data of the bibliographic record.
5. The internal and external staff of MOKKA edits the ‘skeleton’ authority records and the links.

This process is just at the starting point. However, it is considered by MOKKA to be one of the most important tasks to improve the results of searches and to standardize access points for retrieval in MOKKA and, through MOKKA, in the member libraries. Because of many inconsistencies in the catalogs of member libraries, this is a huge, but nevertheless important, task. To help and accelerate this process, MOKKA acquired the Library of Congress Name authorities file and is eagerly awaiting the preparation of the authority files of the Széchényi National Library, to be based on the existing index files and their cross references.

## 5 Subject Approach

Views concerning the role of union catalogs for subject searches are highly variable. One extreme opinion considers the union catalog merely as a tool for finding document data, about the existence (and subject) of which customers are clearly informed. This means that their aim is only to find the library that is able to deliver the given item. The background of this view is that the real tools for subject searches are not the library catalogs at all, but subject bibliographies, citations, etc., and so the task of a union catalog is only the delivery of the document, although the existing retrieval access points (e.g. title keywords or subject headings and classification indices) can obviously be used by the customer.

An other argument against attributing great importance to the subject approach in shared cataloging systems or union catalogs is that in most cases (at least in Hungary and many other countries similar by size and by tradition), the various different subject heading ‘systems’ (if they really are systems, and not merely natural language keywords used as subject headings) prevent the establishment of a consistent, common subject heading vocabulary. The same can also be true for classifications, although some classification systems, like Universal Decimal Classification (UDC),

are widely used or even standardized in many countries of Central and Eastern Europe and are used by most potential members or record suppliers of a shared cataloging/union catalog system. However, there are libraries that do not use classification schemes or use different ones, and even if they use the same system they are often use different updated versions of the system.

On the level of MARC fields, values belonging to different types of vocabularies or schemes for designating subjects can be represented and specified by indicators and/or by subfields. However, retrieval is only possible via the relevant indexes, and in the CORVINA system and in many other systems, there is only one common index for subject headings and keywords and one other for all classifications used. This means that from the point of view of subject search techniques, MOKKA cannot offer solutions for the use of individual subject indication languages.

In spite of all the weaknesses of carrying out subject searches in MOKKA, it is nevertheless possible to use it for that purpose. There are plans for improving the existing procedures, among others by the use of an all-subject thesaurus as a kind of authority file, which can offer a link from various terms to others and can be used for the retrieval of a given subject.

As mentioned above, classification indices and subject headings recorded in the relevant fields and indices of MOKKA are not only those supplied by the library which uploaded the given record, but also include classification indices and subject headings in uploaded duplicate records. As a result, information different from that recorded earlier is added to the relevant fields of existing records, and through this the recall ratio (and of course also the noise) may be increased.

There are also other approaches to subject designation in shared catalog databases. For details, see work by Klára Koltay.<sup>4</sup>

---

<sup>4</sup> Klára Koltay, "Az ODR adatbázis új szolgáltatásai" (New services from the database of the National Document Delivery System), *Tudományos és Műszaki Tájékoztatás*. (48) 2001, No. 8: 315–132.

## 6 Financial Considerations

It has been clear from the very beginning that MOKKA could not be a self-supporting system. Some financial principles have already been agreed upon, whereas other principles and rules will be fixed only after MOKKA has functioned normally for some time (probably at the end of 2002). It has also been finally decided that the record-supplying member libraries will not get any payment for their records. On the other hand, member libraries can download records free of charge. It has not yet been decided under what conditions other libraries can download records. There are two contradictory views about these conditions. According to one view, the system was developed and is maintained first and foremost from national and international resources, and it is accordingly not justified to demand payment for the downloading of its records. Those who support the idea of payment for supplying records refer to the cataloging expenses spent by the record-originating library.

## 7 Development Trends

The most important development tasks are described below:

1. Errors and mistakes detected by internal examination of the system, by the staff of MOKKA, and, last but not least, by the end-users of the system, should be eliminated;
2. The editing of existing (real and 'skeleton') authority records should be started, and this should become a regular maintenance task;
3. Plans for the expansion of the system should move in the following directions:
  - Libraries now outside MOKKA but having a special importance for inter-library loans (public libraries of the counties, further academic libraries and some research libraries) should be invited to join MOKKA as member libraries;
  - Links and direct access to electronic union catalogs for kinds of documents not included in MOKKA (primarily, but not exclusively, serials) should be established, and also the establishment of

interconnected union catalogs for specific types of documents should be encouraged;

- An electronic inter-library loan system should be created within MOKKA, enabling the users to send inter-library loan requests immediately after the identification of the library where the requested document is located and available (MOKKA already offers a link via the Web to the electronic catalogs, holdings data and circulation modules of the library systems used in the member libraries). This enables the user to find the holdings data and the circulation status of the document to be requested); and
- Links to existing virtual union catalogs should be created.

## 8 Virtual or Physical Union Catalogs

The idea of creating virtual union catalogs emerged more than ten years ago. At the very beginning, this was only possible for libraries using the same electronic library system. With the advent of the Z39.50 standard, this possibility became, in principle, a reality for any group of libraries. Nowadays, Z39.50 gateways and other—usually Z 39.50-related—software solutions (METALIB, LibriVision etc.) offer further possibilities for searching in the databases of many libraries by using a single user interface.

Obviously, these technical solutions offer possibilities for establishing virtual union catalogs. However, one could also speak of virtual shared cataloging systems if the system not only searches, but also downloads and—in the case of libraries using different library systems—enables the conversion of records.

The question that emerges from the above technical possibilities is whether, and to what extent, virtual union catalogs can replace the physical (real) ones. It seems that it is easy to answer this question if we reduce the function of a union catalog to executing simultaneous searches in catalogs of various libraries. While it is worthwhile to discuss this question, it must be made clear at the outset that virtual solutions have a lot to offer in comparison with a situation without union catalogs.

Before investigating this question, we quote a paragraph from the executive summary of the Feasibility Study for a National Union Catalog (in the United Kingdom):<sup>5</sup>

Moving from vendor systems to a comparison of physical and virtual catalogs, it was evident in all cases that the physical catalog architecture offered a more reliable, faster and consistent response than any of the virtual systems tested. Comparison of identical searches confirmed the supremacy of the physical model at present, particularly in relation to the user requirements identified in both the conceptual model and the questionnaire survey: for all possible search points the physical catalog showed superior consistency and performance every time ...

It would be easy to close the discussion about real (physical) versus virtual catalogs by referring to the experience gathered by the authors of the above-mentioned Feasibility Study via questionnaires and experiments. However, one could object that the cited opinion is based on a situation in which well-developed physical union catalogs were compared with less developed virtual catalogs. It seems that a further analysis of the possibilities offered by the two solutions is justified.

Let us start with the most important question. A physical union catalog can exist only if it applies a high degree of standardization. One document is represented in the physical union catalog by one single record (in the case of MOKKA there exists a ranking of libraries based on the quality of their catalogs, and if the duplication check finds a duplicate, the record of the higher-ranked library is always kept). The catalog data of other libraries are represented only by the identification data of these libraries. In the case of a virtual union catalog, many slightly or substantially different catalog records of the same document are the result of the search. This means that the physical union catalog offers the same information as the virtual one, but in a uniform way, whereas the use of data available through multiple hits in virtual catalogs can impair the quality and compatibility of catalogs.

---

<sup>5</sup> Peter Stubley, Rob Bull and Tony Kidd, *Feasibility Study for a National Union Catalog. Executive Summary*. April, 2001. <http://www.uknuc.shef.ac.uk/NUCrep.pdf>: 6.

Another important milestone of standardization is the existence of authority files. Their establishment and maintenance is completely impossible in the case of virtual union catalogs. Of course, this is not an easy task in physical union catalogs either, but it can be managed. Without the existence and use of authority control, search results can also have a high noise ratio, but it is even more important that information loss can be very high.

Nevertheless, virtual catalogs can offer the possibility for simultaneous searches in the catalogs of many libraries, in spite of all the problems mentioned above. It is also possible to organize an electronic system for inter-library loans from all libraries, the catalog data of which are available through the virtual union catalog. There is also the possibility to copy retrieved catalog records for cataloging purposes. Taking into account all these possibilities, it cannot be denied that virtual union catalogs can fulfill the functions of shared cataloging as effectively as those of union catalogs. It is also possible to use the solutions offered by the software tools for virtual union catalogs to build the links between various physical union catalogs and/or between physical and virtual union catalogs. It is also obvious that a physical union catalog requires much more effort, manpower, and financial resources, and that a virtual national union catalog or a virtual catalog of any group of libraries offers much more than the searches in scattered electronic catalogs can. However, it should be stated unambiguously that the price to be 'paid' because of the lower quality of virtual union catalogs is too high for a 'core' national union catalog.

## Part 5

### Baltic Union Catalogs



# Chapter 18

## Using a Shared Cataloging System: The Estonian Approach

Janne Andresoo<sup>1</sup> and Riin Olonen<sup>2</sup>

### 1 Introduction

In this paper, we shall focus on the various aspects of designing and implementing a shared cataloging system in the ELNET Consortium's member libraries. We shall try to highlight the joys and sorrows we have faced, and to answer the question whether there is anything we would like to do differently if we could start all over again.

As we both have work experience in the National Library, most of the examples in this paper will be drawn from it.

### 2 Implementation of the System

#### Background

The Estonian Libraries Network Consortium was established on April 4, 1996, and on June 9, 1997, a contract was signed by the ELNET Consortium and Innovative Interfaces, Inc. (III, a U.S. vendor) to implement the

---

<sup>1</sup> Coordinator of Bibliographic Activities, National Library of Estonia; also Team Leader of Cataloging Working Group, ELNET Consortium, [janne@nlib.ee](mailto:janne@nlib.ee).

<sup>2</sup> Library System Manager, National Library of Estonia; also Executive Manager, ELNET Consortium, [riin@nlib.ee](mailto:riin@nlib.ee).

integrated library system INNOPAC in Estonia. The integrated library system INNOPAC was officially presented at the 7th congress of Estonian librarians on October 22, 1998, and the shared system went live on January 1, 1999 (after one and a half years of testing, adapting and intensive user training). Currently, the libraries of the ELNET Consortium are moving to the new Web-based system Millennium.

The implementation of the new library system inevitably brings many changes in people's everyday work—new tasks and different responsibilities for staff, changes in work routines and because of that, reorganization of the library's workflow. And the larger the library, the larger the number of possible changes and the larger the staff that will inevitably have to adapt to those changes.

### IT-specific training

The time period during which the library system was implemented also brought many changes to the National Library of Estonia. The use of information technology in general has become broader (even in those workplaces which had not previously been automated until now). To provide the entire library staff with a basic knowledge of computers, several in-house training sessions and outside courses were arranged. To extend and improve training services, a computer class was organized in the National Library, later serving as a training base for all member libraries of the ELNET Consortium, which equipped it in part.

### New Rules for Data Input

New rules were applied to the data input—paper-based bibliographic descriptions or those in older databases were totally different from the data input in INNOPAC, which uses the MARC21 format for handling and saving data. Since the open system INNOPAC makes new demands for the unification of the data input, the acceptance of unified standards has become vitally important. Besides following international standards, it has become very important for librarians to agree and compromise on the national level. The importance of data quality and standards was also pointed out by Bohdana Stoklasová of the National Library of the Czech

Republic at the conference “National Bibliography in a Changing Information Environment,” organized by the National Library of Estonia in 2000.

Despite the fact that the library staff was not familiar with MARC21, the principles of machine-readable cataloging were not completely new to them. In 1993, the National Library purchased a Finnish integrated library system KIRI. For data input and storage, KIRI used the FINMARC format (the Finnish version of the MARC format). During the implementation period it was hoped that this system could be developed and adapted to meet the Estonian research libraries’ needs. Unfortunately KIRI did not satisfy these expectations. Still, the attempts to put KIRI into operation were not a complete waste of time, for the staff had a chance to become acquainted with the rules of machine-readable cataloging, to perceive the principles of an integrated library system and its impact on everyday work, and to get enough information to be able to evaluate library systems better next time.

In addition to the MARC21 format, Anglo-American cataloguing rules (AACR2) and the principles of authority control were applied in the cataloguing process (ISBDs were already in use in Estonia). In addition, the principles of copy cataloging were also new to us. Our specialists had to learn first themselves and demonstrate later to the rest of the personnel how the Z39.50 protocol worked. Subject indexing was another challenge for us. The staff had no problems with classification (Estonian libraries use the UDC), but they were not so familiar with subject indexing, for we had not used it earlier to any great extent (the Estonian Universal Thesaurus was not published until January 1999).

The unification of data input has also become very important for us, since we share our database with other member libraries of the ELNET Consortium. To improve cooperation, we formed several special working groups (for instance, for cataloging, serials processing, authority control, system management questions, etc). The mandate of these working groups is to establish and revise professional rules, standards and detailed operating regulations, organize training activities and draw up the criteria for quality evaluation, etc. In these working groups, specialists from different libraries discuss their problems in order to find common solutions.

## Changes in organization

In addition to the changes mentioned above, it was necessary to optimize and adapt working routines to the demands of the new system, and because of that there was a need to reassess the entire library's workflow. The reassessment of the library's working routines was the task of an implementation group that included all directors and key specialists. This reassessment changed the overall work organization of the National Library: some existing departments had to be integrated, and some new ones had to be established. For instance, three new structural units were established: the Authority Control Department, the Retrospective Conversion Department and the Re-Cataloging Department.

## Training courses

First, catalogers received training. In the spring of 1997, the National Library invited colleagues from the Helsinki University Library to introduce the MARC format and to provide basic training for representatives of all ELNET Consortium's member libraries. This training was supported by the NORDINFO. At the end of the same year, we had the opportunity to meet Sherry K. Little from Texas, USA, who shared with us her knowledge and experience of USMARC and authority control.

Training for using the new system began gradually at the end of 1997. First, specialists who were (and are) responsible for the further training of the staff had their training sessions. During the year and a half ending in January, 1999, when the system was launched, four training sessions were arranged in Tallinn and Tartu by a representative of Innovative Interfaces, Inc., and countless other training sessions were organized by our own specialists. This was a very intensive period of time for these key persons, since they had to be prepared to start training others immediately upon completion of their own training. By now, some additional support persons (specialists) have been trained in the main modules of INNOPAC in almost every library.

The cataloging staff was the most active group in the testing and training phase, and it was the best-prepared to start working with the new system. The most averse to the system was the acquisition staff, first,

because its training started somewhat later, which meant that there was less time to practice and get familiar with the system, but probably for the main reason that the acquisition staff had to change its work procedures more than others.

### Tour de INNOPAC

Planned rearrangements had to be tested against the reality. To be absolutely confident about the decisions we made, to discover all possible mistakes and contradictions within the newly planned work routines, to test the results of training and to identify deficiencies in that area, and to make people understand what it meant to be working with the new system, an expert group organized a simulation of real work, or Tour de INNOPAC as we also called it. It meant that we actually recreated all work procedures, ranging from placing an order to placing a received and processed item on the shelf, and we simulated that process separately for books, serials, printed music, etc. The actual library staff was involved in its real work environment, and we actually accompanied the items through all these procedures (through acquisition, cataloging, subject indexing, authority control, etc) and through all the departments which were involved.

### Further plans

Although the system has been in operation since January 1999, the training process is not yet complete. We still need advanced training courses, with training associated with the changes in the automated library system (for instance, right now we are moving to the new Web-based version of our library system, Millennium). Likewise, we still need to train new staff (not only at the libraries themselves, but also at the Department of Information Studies of the Tallinn Pedagogical University and the Viljandi College of Culture).

In dealing with training and the introduction of a new library system, one should not forget the users. There will always be a great need for training library visitors. One may argue that it is not necessary to train users, because INNOPAC is so easy to use. While that is true, it is also the case that we have entered some very important agreements in the ELNET

Consortium that also (directly or indirectly) affect searching in our electronic catalog. It is clearly very important for our patrons to know the relevant details.

### 3 Structure of the Shared Catalog(ing) System

The attribute that most accurately characterises the Estonian common cataloging system is ‘multifunctionality.’ The main functions are:

1. Union catalog (including retrospective conversion);
2. National bibliography database;
3. Database of CIP records;
4. Database of articles.

#### Union Catalog

The idea of union catalogs in Estonia has been related, above all, to providing information about foreign acquisitions (books and serials) in research libraries. Starting as a card catalog (books since late 1950s and serials since early 1960s), the union catalogs were also published in book format until 1997/1998. There were many reasons for concentrating on foreign material; because it was hardly possible to purchase foreign literature during the Soviet period, the union catalogs served as a basis for coordination of foreign acquisitions. Over 30 libraries were involved in this cooperation.

Discussions about a national union catalog, which could provide information about all holdings, at least in larger libraries in Estonia, started in the 1990s. There was a plan to establish a common information system which would be based on (preferably) one integrated library system and, in the beginning, involve 10 research libraries acting as cataloging centers, and 2 main regional public libraries. One can see a realisation of this plan in the Estonian Libraries Network (ELNET) Consortium, which was established in 1996 by seven research libraries and by now includes 13 libraries (also 2 main public libraries).

In an electronic environment, a union catalog could be organized in three ways:

1. As a centralized system with a central database;
2. As a clustered system in which a group of regional libraries support a single database;
3. As a decentralized system in which each library maintains its own database, but a union catalog is shared by all libraries.

The first plan of the ELNET Consortium was to implement the first model, a centralized system with just one database for all its member libraries. Even a name was chosen for the database—ESTER, which is a combination of MARC codes for the Estonian language (EST) and country (ER). Besides that, the word is just a beautiful female name. For a number of reasons,<sup>3</sup> we chose the second model, a clustered system with two regional databases, one in Tallinn (<http://helios.nlib.ee>) and one in Tartu (<http://merihobu.utlib.ee>). Even though we have two separate databases, we emphasize that we still have only one system based on common principles. We gave both databases the same name, ESTER, supplemented by the name of the city—Tallinn or Tartu.

Górny and Nikisch have pointed out the benefits and deficiencies of different types of organizational and technological structures in creating a union catalog.<sup>4</sup> These arguments are correct in a situation where each participating library maintains its own database, and the objective is to create a separate database—a union catalog. Then, indeed, establishing a centralized catalog implies higher costs for the construction and maintenance of a catalog, and a virtual union catalog may be cheaper to build and maintain. In our situation, where all participating libraries had to

---

<sup>3</sup> Asko Tamme, “A Shared Catalog in Estonia: Do We Need a Union Catalog?” Paper presented at The Andrew W. Mellon Foundation Conference on Union Catalogs, Tallinn, October 17–19, 2002. <http://www.nlib.ee/inglise/docs/mellon/melprog.html>.

<sup>4</sup> M. Górny and A. Nikisch, “Evolution of Union Catalogs—the Union Catalog in the Digital Age.” Paper presented at the Union Catalog Conference in Tallinn, October 17–19, 2002; also at <http://www.nlib.ee/inglise/docs/mellon/melprog.html>.

purchase and implement an integrated library system first, the shared catalog environment was more rational. Instead of purchasing server and library system for each individual library and implementing systems separately, we chose to have just two servers, two installations, and hence two catalogs.

### Starting-Point

In 1999, we did not start with empty catalogs. We had several important databases that could be folded into the new environment. In Tallinn, where four libraries (National Library of Estonia, Estonian Academic Library, Tallinn Pedagogical University Library and Tallinn Technical University Library) shared the same catalog, it was mainly the databases from the National Library that formed the starting-point for the new electronic catalog. They were two national bibliography databases: books published in 1991–1998 (approx. 21,500 titles) and serials published in 1994–1998 (approx. 1,350 titles), and two main databases of foreign materials: the union catalog of foreign books—books received by larger research libraries in 1993–1998 (approx. 53,000 titles, supplemented with approx. 39,500 additional titles from the database of foreign acquisitions), and the union catalog of foreign serials—serials received by larger research libraries in 1993–1998 (approx. 12,500 titles). In addition to these, the circulation database was also converted (approx. 51,000 additional titles, 117,500 patrons and 75,000 check-outs). In Tartu, where three libraries (Tartu University Library, Estonian Agricultural University Library and Archival Library of the Literary Museum) shared the catalog, the starting-point was the previous electronic catalog of the Tartu University Library, INGRID (approx. 40,000 titles).

Since 1999, during the years that we have been using INNOPAC, we have converted some additional databases from different environments (for instance, Estonian serials published in 1766–1940, Estonian maps published in 1988–1998, etc.). Other libraries that joined the ELNET Consortium (two university libraries: the library of the Academy of Arts and the library of the Academy of Music, both in Tallinn; and two main public libraries: the Tallinn Public Library and the Tartu Public Library) came with their previous electronic catalogs. On the one hand, the

conversion to ESTER was very useful for the new member libraries, but on the other hand, each conversion created minor chaos in the database. Every possible method of identifying duplicates was applied before conversion, but it still created problems.

### Retrospective Conversion

The implementation of INNOPAC led to a number of development projects, in particular retrospective conversion projects. In 1998, the ELNET Consortium received two grants, one from The Andrew W. Mellon Foundation (\$165,000) and one from the Open Estonia Foundation (through the library program of the Open Society Institute, \$100,000). At the beginning of 1999, the third grant was obtained from the Cultural Endowment of Estonia (78,000 EEK). Following the corresponding experience of other countries, Estonian libraries started the retroconversion of national bibliography data. The plan was to manually enter the data from the card catalogs or the published national bibliography and other bibliographies into the database. The main criteria for selecting a method of retrospective conversion were cost and quality; or actually, the best compromise between cost and quality. We had to take into account that there have been several changes in cataloging rules in Estonia, and our catalogs also included many old and handwritten cards. It was also not possible to copy records of Estonian publications on the national bibliography level. The decision was to key records manually, involving professional catalogers and also trained non-professionals in the process.

The coordinators of these projects were from the National Library of Estonia and the Estonian Academic Library, since these two libraries have been sharing the responsibility for compiling the national bibliography according to the Estonian retrospective national bibliography program launched in 1978, which covered Estonian books from 1525 to 1945 and serials from 1675 to 1945. The National Library is responsible for the period from 1945 to the present, and the Estonian Academic Library for the period before 1945. In the retroconversion projects, the bibliographic descriptions were provided by the coordinating libraries, to which each library added information about its holdings.

In 1999, three retroconversion projects were launched (coordinated by the National Library of Estonia):

1. Retrospective conversion of Estonian books published in 1945–1991. This project is almost completed by now; approximately 107,000 titles and 800,000 items have been entered and besides books also approx. 4,400 titles and 12,800 items of printed music have been entered in the database. Currently there are only some tests to be done to check the completeness and quality of data.
2. Re-cataloging of books in Estonian published in 1918–1940 (approximately 24,000 titles, almost 80% of the total, and 92,500 items have been entered).
3. Re-cataloging of Estonian periodicals published in 1945–1993 (approximately 2,600 titles, more than 80% of total, and 125,400 annual sets have been entered).

In 2002, one additional retroconversion project was launched (coordinated by the Estonian Academic Library):

4. Re-cataloging of books in Estonian published before 1917 (approximately 12,250 titles, almost 70% of the total, have been entered).

The Estonian Academic Library was also responsible for the retroconversion of Estonian serials (excl. serials in Russian) and Estica (books and serials in Estonian published abroad). Both are almost completed by now.

Although the priority of retroconversion was the national bibliography, the libraries have been doing a huge job in the retroconversion of foreign material. Depending on the time and place of publication, libraries are trying to find sources for copy cataloging, but there is still sometimes a need for original cataloging. During the last two years, the ELNET Consortium has found some resources to also support the retroconversion of foreign material. With this small level of support, approx. 147,500 titles and 162,700 items have been entered in the database.

The major deficiency in retroconversion projects is the lack of subject indexing. A decision was taken not to have subject indexing together with descriptive cataloging, because the catalogs and bibliographies did not have subject headings earlier and the retroconversion was not carried out *de visu*. The second reason was that subject indexing could have slowed down the

whole process considerably; furthermore, we did not have enough skilled staff for that extra job.

## The Current Situation

Currently, 13 libraries (nine in Tallinn and four in Tartu) are participating in supplementing the catalog ESTER. The last two that joined were the Deposit Library of Estonia and the Medical Library of Estonia, both in Tallinn. All member libraries use the same database for their everyday work (or actually, the libraries in Tartu use one database and the libraries in Tallinn use the other one). This means that each title is cataloged (or copied from another catalog) only once, by the library that receives it first, and every additional library simply attaches its holdings to the title. Where Estonian publications are concerned, the libraries (other than the National Library) may enter only mandatory data about the title, as the National Library is responsible for cataloging Estonian publications at a national bibliography level. All other materials need to be fully cataloged by the first receiving library.

The catalog ESTER Tallinn contains approx. 650,000 titles of multi-language materials owned by seven libraries in Tallinn (the data from the Deposit Library of Estonia and the Medical Library of Estonia are not converted yet) and the catalog ESTER Tartu contains approx. 500,000 titles of multi-language materials owned by four libraries in Tartu. The titles identify books (approx. 80% of the entire database), serials, periodicals, maps, printed music, videos, sound recordings, offline and online documents, etc., owned by these libraries. Approximately 20% of these titles can be found in more than one library. At present, due to the intensive retroconversion of Estonian publications, the major part of records is in Estonian (approx. 30%) followed by records in Russian, English, and other languages. Approximately 38% of titles are published in Estonia. Since the retroconversion projects are going to be completed very soon, and in view of the fact that it is mostly research libraries that participate in the creation of the union catalog, it is safe to predict that the percentage of records describing foreign material will increase rapidly.

In addition to the titles, the catalog also lists each copy of these titles (approx. 2,100,000 in Tallinn and 1,350,000 in Tartu; but this amounts to

only approximately 26% of all holdings actually owned by participating libraries). And since we have a shared catalog, with all data in one catalog and no separate union catalog, it also tells the user right away how many copies we have and in which library, whether the copy is in the library and available for checkout or whether it is already checked out, or just ordered for the library.

The library user can search either in the whole database (separately in Tallinn or Tartu) or just in one virtual part (scope) of the database. It was decided to provide scopes by the level of description (monographs, serials, analytical records) and by the individual libraries. If users cannot find the required information in one (Tallinn or Tartu) database, they can direct their search to the other system very easily, as these systems work as partners. Even though this redirection is very easy to do, with just one keystroke, many users complain about it. Because of this and for other reasons, there have been serious discussions during the past years about integrating these two catalogs. Lately, more support has been given to the idea of not joining databases, but leaving them separate and providing a common search interface for the users. And since these catalogs incorporate the same principles and the same indexes, with the same indexing rules, we will not face the problems arising from broadcast search for our two regional union catalogs, which were described by K.Coyle in her paper on the MELVYL union catalog.<sup>5</sup>

Even though all member libraries use the same system and the same database, they do not offer all the same services to their users. For instance, only the National Library, the Estonian Academic Library and the Tartu University Library offer their users the ability to put a hold on items they require. The only library still maintaining its own local electronic catalog (based on entries from the shared catalog) is the Tallinn Technical University Library.<sup>6</sup> On the whole, a great deal remains to be done in the libraries. Two more serious problems that the ELNET Consortium is facing in providing services to users are sorting the search results in the keyword

---

<sup>5</sup> Karen Coyle, "The Virtual Union Catalog," in this volume.

<sup>6</sup> Jüri Järs, "An Interface Between Union Catalog and Local Catalog: Why and How?" Paper presented at The Andrew W. Mellon Foundation Conference on Union Catalogs, Tallinn, October 17-19, 2002. <http://www.nlib.ee/inglise/docs/mellon/melprog.html>.

index (search results are not displayed in the correct alphabetical order), and the display of Cyrillic characters (at present users cannot see records in Cyrillic in the Web catalog). Lack of user manuals, guidelines or simple instructions form another source of problems.

Notable achievements include some special software compiled by the ELNET Consortium for the public libraries in Estonia. Today, most public libraries in Estonia are using a Finnish integrated library system, Kirjasto3000, which uses FINMARC format and does not support the Z39.50 protocol. Because of that, the libraries cannot copy records from ESTER very easily. To provide a better service for public libraries, the ELNET Consortium has created a special converter, US-FIN, for them.

### National Bibliography Database

In addition to the shared electronic catalog, ESTER also functions as the national bibliography database. Out of the total number of Estonian publications, comprising books published since 1525 and serials since 1675 (Estonian-language serials since 1766), approximately 80% are already included in ESTER.

Currently, the national bibliography database is an integral part of the union catalog and is not even scoped separately. It is also a problem that in a shared system environment, the national bibliography records are not adequately protected from accidental updating (the agreements in place only stipulate that certain data will not be changed in these records after certain libraries have declared them definitive). Therefore the National Library has been developing a separate database environment for the national bibliography database.

Since the beginning of 2002, the legal basis for compiling the national bibliography has also changed. In March 2002, the amended National Library of Estonia Act became effective, and under this Act the national bibliography database has acquired the status of a state database. This new regulation gives a new legal meaning to the database, increasing the responsibility of the authorized processor of the database and stipulating stricter requirements for data protection.

Since the data input in the ESTER database follows all important international standards and national agreements, the future national

bibliography database will be based on records entered in the ESTER database. This will also guarantee the uniformity of data in both databases.

Besides having a separate database, the possibilities of combining the national bibliography database with other databases will represent added value for the users. For instance, the use of the database of statistics for Estonian print output and the databases of national ISBN and ISSN centers would provide more information about Estonian publishers and printing houses, together with their publications.

### Database of CIP Records

Beside the shared union catalog and national bibliography data, the ESTER Tallinn database also contains CIP records (cataloging-in-publication). Currently, there are approximately 700 CIP records in the database.

The implementation of INNOPAC created a new basis for cooperation with publishers and booksellers. One form of cooperation between publishers and libraries is the national cataloging-in-publication (CIP) program, which enables publishers to inform the national bibliography agency about books that are going to be published. The ISBN and ISSN Centers operating in the National Library of Estonia started CIP cataloging at the beginning of 2000. By granting an ISBN to a monograph or an ISSN to a series, the Centers obtain comprehensive information on the given publication from the publisher that is entered in the electronic catalog of the ELNET Consortium. CIP records serve as a basis for a definitive record at national bibliography level. A record is excluded from the database of CIP records (by changing the status of the record) after the publication arrives in the library. For libraries, CIP records help to monitor the supply of legal deposit copies, and for publishers they act as advertisements because this information reaches all interested parties through the electronic catalog.

### Database of Articles

Last but not least, ESTER also functions as a basis for joint efforts to create a database of articles from Estonian serials (excluding newspapers).

Numerous discussions among Estonian research and public libraries led to a decision in 1998 to discontinue the publication of the bibliography of

articles from Estonian periodicals in printed form. Instead, libraries decided to cooperate in creating analytical bibliographies, and to include records of articles in the ESTER database. The database of Estonian articles contains approximately 50,000 records from 180 titles of serials, and grows by about 14,000 records annually. The database of articles is the only part of ESTER that is mirrored in both systems. New records are copied from one system to another twice a month. This cooperation involves six member libraries of the ELNET Consortium. Titles are divided among participants by subject according to the libraries' profiles.

Last year, some libraries had problems in living up to previous agreements, in particular regarding the speed of updates, and because of that there have been serious discussions about the future prospects of this cooperation.

### Further Plans for Cooperation; Development Projects

The implementation of INNOPAC provided an opportunity to start a number of development projects, mostly related to online publications and digital library programs. This paper describes only those projects that are related to the union catalog.

From the mid-1990s, everybody could observe a tremendous increase in publishing on the Web. To help library users to orient themselves in this world, several libraries started to collect information about valuable Web-resources into subject gateways. It has also become clear that in the Web environment, publications have a tendency to disappear from the Internet. So the National Library, being responsible for preserving the national cultural heritage, launched the project ERICA (Estonian Resources on the Internet: Cataloguing and Archiving) in March 2000. The aim of the project is to work out methods and means for collecting, registering and making available Estonian online publications. The elaboration of the selection criteria for online publications was started on the basis of their registration in the national bibliography. A positive decision about a given online publication results in the addition of a MARC record to ESTER, and then also to the national bibliography database. By May 1, 2002, the National Library had identified, collected and systematized in thematic lists approximately 500 Estonian online monographs and 400 periodicals (with

the domain .ee). Only a small number of these (periodicals with ISSN) have been entered in the ESTER database. Due to the fact that collecting, cataloging and preserving online publications requires considerable additional resources, i.e. extra time, money and staff, the Consortium's member libraries initiated discussions about cooperating in the creation of an Estonian virtual library (the leader of this project is from the Pedagogical University Library).

Along with the project ERICA, the National Library has gained valuable experience in collecting and preserving electronic publications in the course of the pilot project ARES (Electronic System of Articles). The purpose of the project was to work out the technology and principles of how to collect, preserve, and then provide access to, full texts directly from the electronic catalog ESTER (via the corresponding fields in MARC records). The project covered materials protected by copyright, and thus required corresponding agreements with publishers and authors. Due to the lack of sufficient resources, the project was stopped.

#### 4 Conclusions

What conclusions can we draw from the previous years?

**The attitudes of the people towards the implementation of the system must never be underestimated, because it can make the difference between success and failure.**

People are naturally averse to change, especially when the changes involve new technology. To minimize the negative side effects of innovation, to help staff to accept forthcoming changes and not to oppose them, it is very important to involve them from the very start, to keep everyone informed of progress at regular intervals (and what is important, not to cover up mistakes), to introduce further plans, to maintain a positive attitude, and to reassure individuals about their importance in the implementation of the library system. And this not only during the implementation period, but also later, when the system is used routinely.

Flexible and collegial management of cooperation was one key to the success in Estonia.

The implementation of the new library system in Estonia has been extremely successful, particularly the first years. We achieved a lot within just a couple of years: a new organization, the Estonian Libraries Network Consortium, was established, a new system chosen and implemented, several important retrospective conversion projects started and carried out, etc. All participating libraries could take part in decision-making, and beside formal, recorded meetings there have been many informal meetings. Instead of fighting with bureaucracy, experts were dealing with substantive questions. Now that the number of member libraries has increased, the number of tasks in the consortium has also increased (besides a shared system and a union catalog, there are new topics such as the licensing of electronic publications and digitization). Therefore the role of administrative management has become more important, and we need more written contracts and agreements than before. Especially now that the initial funding for implementation and retroconversion is running out, and we have to carry on with our own resources.

The importance of documentation cannot be underestimated.

The most serious drawback that we are facing now, as a result of the rapid development during the first few years, is deficient documentation. There have been so many decisions to make and problems to solve that there was not enough time to write it all down. To some extent this applies even today.

The implementation of a shared system represents the only possibility for us to have a good library system that is highly valued worldwide. But the decision to put all information into one database does not seem to have been the best decision.

Estonia is a small country with limited resources, and cooperation has been the only way to obtain good software. The decision to put different types of data into a single database was mostly motivated by the desire to save on resources. The idea was to allow users to make just one search to find different types of information. Libraries do not need to maintain different databases and environments, each record has to be keyed only once, and

with limited resources and overlapping subjects it is important to know what other libraries already have. And users do not need to remember different addresses and search interfaces. On the other hand, it seems that amalgamating everything has made the general overview less precise and has confused some users. The only reasonable answer seems to be training, user guides and instruction.

And finally, at the beginning of this paper we promised to answer the question whether we would like to do anything differently if we could start all over again. The answer is: not very much. We personally feel that we should have paid more attention to documentation and hurried less with the development.

## Part 6

### South African Union Catalogs



# Chapter 19

## A National Union Catalog for Shared Cataloging and Resource Sharing by Southern African Libraries

Pierre Malan

### 1 The Founding of SABINET

In 1979, the South African National Library Advisory Council (NLAC) initiated a national project to investigate the feasibility of establishing a library network and national union catalog, the South African Library Network (SALNET). The groundwork for this project, also known as the Computerized Cataloging Network Project (CCNP) was laid by the former MARC Working Group of the NLAC, which already started feasibility studies as early as 1970.<sup>1</sup> The MARC working group was also responsible for the development of SAMARC (South African MARC) based on UNIMARC at that time,<sup>2</sup> which set a standard that would have a great impact on future developments.

Recommendations made by NLAC indicated that there was consensus among libraries in South Africa for the establishment of SALNET. The main purpose of the establishment of the network would be to facilitate

---

<sup>1</sup> M. C. Boshoff and H. F. Bester, *South African Library Network (SALNET), Functions and Network Architecture*, 1980.

<sup>2</sup> National Library Advisory Council, Committee for Computerized Cataloging Network, Working Group for Bibliographic Standards, *SAMARC* (Pretoria: NLAC, 1980).

resource sharing among South African libraries, mainly by allowing shared cataloging and an inter-library loan mailbox service.

Certain principles were established for the creation of the network. These principles were not only very significant at that time, but are still applicable today. It also turned out to be the case that significant problems emerged when there were deviations from these principles. These principles were:

1. The system should be as simple as possible within the framework of a networked central library system;
2. Participation in the network should be cost-effective for libraries;
3. The purpose of the system should be to serve the user and not only the librarian;
4. The autonomy of local library systems and computer centers should always be taken into account;
5. The system should lend itself to the creation of a central database with high integrity; and
6. The central database should provide good coverage of materials in participating libraries.

The recommendations were presented to the Department of National Education by NLAC and were accepted by Government in 1981. SABINET (originally referred to as SALNET) was officially constituted on February 28, 1983,<sup>3</sup> when forty-six libraries and information centers made a ten-year commitment to establish the network.

## 2 The Start of Computerization

Before the South African Bibliographic and Information Network (SABINET) was founded, an extensive study had been conducted, and the SABINET project team had decided to use the program package of the

---

<sup>3</sup> F. W. Coetzee, *Historical Overview, South African Bibliographic and Information Network, Annual Report 1984 to 1985*.

Washington Library Network (WLN), later also known as the Western Library Network, from the United States on an interim basis for the SABINET system. The WLN programs, including their bibliographic database, were installed and accessible in South Africa as of September 24, 1983.

SABINET contracted with a service bureau, Automated Business Systems (ABS), for the provision of computer facilities, and access to the services was through an established national government IBM SNA network called GOVNET.

The State Library (now the National Library of South Africa) was the first member to be linked to SABINET, followed shortly after by the South African Bureau of Standards and UNISA (University of South Africa). By March 31, 1984, 13 members were linked to SABINET. The only service available was an inquiry function on the 2.7 million record database housed on the WLN system. Within the months that followed, many more members were connected to the network, Library of Congress records were being batch-loaded into the catalog, and the functionality was extended to online cataloging.

Since the decision to use SAMARC as a bibliographic standard had already been taken prior to 1980, it was urgent not only to have the interim WLN system as compatible as possible with SAMARC, but also to develop a full-blown SAMARC system for South Africa. To this end, by March 1985, an interface (SABIMARC) was developed on top of the WLN system, which allowed SAMARC tagging with the existing USMARC punctuation. At the same time, SABINET issued an invitation to tender for the development of a unique SAMARC system to conform to all expectations for a South African union catalog. By the middle of 1995, the board of SABINET had appointed the chosen company to develop and implement the yet-to-be-developed SAMARC system.

The SABINET Managing Director informed the members that the activities planned until the end of 1987<sup>4</sup> would include:

---

<sup>4</sup> F. W. Coetzee, *Planned Activities 1985 to 1987, South African Bibliographic and Information Network Annual Report 1984 to 1985*.

1. Establishment of efficient maintenance services for all implemented functions on the WLN system;
2. Support of the SABIMARC interface and enhancement of the quality of this interface even further; and
3. Progress with the phased development and implementation of the full feature SAMARC bibliographic system. (The full implementation was scheduled for a three-year period.)

Two years after the establishment of SABINET, after just succeeding in putting a working solution in place, the announcement was made that the development and implementation of the SAMARC union catalog system would be completed in only three years. The importance of achieving this goal was underscored by the fact that the implementation and further maintenance of the SABIMARC interface excluded SABINET from implementing any further WLN software upgrades.

### 3 Local Development of the SAMARC Union Catalog System

Datatrust, a local software development house to which the tender was awarded, started the development of the SAMARC system during the second half of 1985. A year after the start of the development of the system, the development team requested that an additional investigation outside the scope of the original project needed to be undertaken, to allow for the detailed investigation and specification of the SAMARC system requirements. This already indicated that there were severe shortcomings in the original specifications on which the project was based.

During the second year of the development, it was reported that approximately 50% of SABINET personnel time was dedicated to the SAMARC system project. Staff involvement ranged from detailed analysis to development of further specifications, programmer support and testing.<sup>5</sup> Although according to original planning, certain modules were to have

---

<sup>5</sup> I. van Niekerk, *Departmental Report: Information Services, South African Bibliographic and Information Network Annual Report 1987 to 1988*.

been available for client implementation at this stage, all were still in various stages of development and testing. Even at this stage, indications were that the project was poorly managed and starting to fall behind schedule.

Indications of very serious difficulties with the project started to surface in the 1988/89 financial year, when it was reported in the SABINET annual report that

1. the development of the new SAMARC system was still dominating all activities of SABINET;
2. that progress was seriously hampered by the resignation of various key staff members; and
3. that delays were incurred with the introduction of database conversion programs, "a task which proved to be more extensive than foreseen during the initial planning stages of the project," according to a report.

At this stage, the newly appointed Managing Director of SABINET, Gerhard Kemp, started to view the status and progress of this inherited project very critically. The following information surfaced after various actions were put in place in an effort to steer the project back on course:

1. It became evident that the project was poorly managed. The software development company involved in the development was too scared of losing the contract or of seeing it end prematurely. This made the company withhold information about the true status and achievability of the project. Furthermore, staff inside SABINET also withheld damning information, knowing that the failure of the project would have an unfavorable impact on their employment. The truth about the poor status of the development only surfaced after the appointment of a new project manager who had nothing to lose in exposing the truth.
2. A further warning sign came from the computer bureau where the WLN system was hosted and on whose platforms the development of the new system was taking place. The bureau indicated that the mere testing of the new system used so much more computing capacity than the existing live WLN system that it would not be in a position to host the new development in a production environment due to lack of capacity.

3. With all information eventually exposed, the SABINET management calculated the cost of completing the development. The calculation showed that to merely complete the developments that were currently under way to a point of usability would cost no less than R 3 million.<sup>6</sup> This figure excluded the tremendous costs that would be involved in finding the necessary computer mainframe infrastructure that would be necessary to cope with the demands of the software application.

The above revelations finally brought home the realization that it was not wise to continue with the systems development. This decision was very unpopular with SABINET staff, and perhaps also among some in the SA library community.

After a thorough investigation done by external consultants during the second half of 1990, all concerns relating to the continuation of the development of a unique SAMARC-based system were confirmed. During November 1990, the Pythia Project (as the system was later called) was finally scrapped. Sadly, this development, with a direct cost to SABINET of nearly \$2.7 million and a total cost of \$10 million, was never to be implemented and nearly resulted in the demise of SABINET and of all prospects of having a National Union Catalog in South Africa.

#### 4 Implementation of ERUDITE

Early in 1991, an emergency SABINET board meeting was held to decide on the future, given the final decision that the Pythia Project would not continue. The only options really open for discussion were either to continue with the WLN system, or to implement an alternative existing library automation solution which complied at least with the SAMARC standard. Although perhaps the easiest solution for SABINET would have been to continue with the WLN system, there were unfortunately many factors that argued against it. Perhaps the most important was that the system, implemented in 1984, was never upgraded because of its custom-built SAMARC interface, and was therefore falling far behind in usability

---

<sup>6</sup> South African Rand: at the time one US dollar was approximately R2.5395.

and functionality according to 1990 standards. The system was also still based on mainframe computing technology, which was becoming increasingly expensive to operate, while cheaper alternatives such as UNIX platforms were starting to become the norm.

During this meeting, it was decided to draw up the system requirements for a new system and to issue a tender for the supply of an alternative system within a period of six weeks. Due to the sanctions still being imposed on South Africa by Western countries at that time, and with SAMARC still a very prominent requirement, it was likely that the preferred vendor would be South African. The process was completed in record time, and after requesting tenders, SABINET received proposals from three local system vendors. The contract was finally awarded to a local company for the implementation of the ERUDITE library system. The system was to be implemented on a UNIX platform, which meant substantial savings in operating costs for SABINET. The total cost of the system, hardware and implementation was less than what it would have cost to complete the development of the Pythia system.

By April 1992, the implementation of the ERUDITE system was completed and the WLN system turned off.<sup>7</sup> This marked a new era for SABINET, with a user-friendly SAMARC-based system that was also accessible through networks other than the GOVNET network. In the following years, the number of users and usage of the service gradually increased. Services were further complemented by the addition of an ILL (Inter-Library Loan) module that was a joint development by SABINET staff and the owners of the ERUDITE system.

When SABINET purchased ERUDITE during 1991, the system was distributed by one of the largest computer companies in SA. However, two years later the division responsible for ERUDITE was sold and has since then changed ownership many times. This unstable ownership situation and the resulting lack of a clear strategy contributed to the fact that the systems developed very little in later years.

While compliance with SAMARC was a very strong motivator for the choice of ERUDITE in the early 1990s, it became a big stumbling block in

---

<sup>7</sup> G. Kemp, Report from the Managing Director, *SABINET Annual Report 1991/1992*: 5.

later years, when the SA Bibliographic Standards Committee decided unanimously in 1998 on the implementation of USMARC in South Africa. With the increasing implementation of USMARC-based systems in the country, and because of the lifting of sanctions, the SAMARC-based National Union Catalog was quickly being outgrown by its USMARC-based members.

Seven years after the implementation of ERUDITE, SABINET was again confronted with many problems which necessitated the migration of the South African Union Catalog (SACat) to an alternative platform.

### Problems with the ERUDITE System

The ERUDITE system on which the SACat was housed needed to be replaced for the following reasons:

1. It was functionally outdated, e.g. keyword searching was slow;
2. Its DBMS (database management system) was technologically outdated;
3. It was not Year 2000-compliant;
4. It was not USMARC-based;
5. It was not designed to handle very large databases;
6. It was resource-intensive in terms of computer hardware infrastructure, which affected the speed of the batch loading of records, thus interfering with SABINET's ability to update and maintain its databases on the network system; and
7. Because of the instability of the vendor and the vendor's lack of capacity, SABINET was left with little or no support.

### Problems with the SACat

SACat struggled with a number of problems:

1. It had no authority control over names and subject headings used, which affected the quality of retrieval;
2. It had bibliographic records of differing quality, which made shared/copy cataloging and searching very difficult; sub-standard records were often those loaded via tapes from user catalogs;

3. It had many duplicates because of poor matching algorithms, so holdings could be attached to multiple records;
4. It was in SAMARC and needed to be converted to USMARC;
5. Its holdings were not always kept up-to-date by member libraries, including academic libraries; and
6. There was little or no machine validation of headings, tagging, etc.

### Problems with the Inter Library Loans System

Of all the SABINET services, the Interlending Module, custom designed for South African circumstances, is the most popular one among users. However, the following problems existed:

1. It was built on ERUDITE and was therefore functionally and technologically outdated;
2. It interfaced with the SACat, and therefore inherited all the SACat problems described above;
3. It made heavy use of hardware resources;
4. It required a high level of support, since it was custom designed by Sabinet Online;
5. It did not pay for itself in terms of usage;
6. It only allowed loans mediated by librarians, and did not permit unmediated end-user lending.

## 5 SABINET and Sabinet Online

In January 1997, a new private company, Sabinet Online (Pty) Limited, was formed with the objective of addressing the changing needs of the online information community and to keep pace with the rapidly changing technology. SABINET's operational activities were sold to Sabinet Online, and a contractual agreement was entered into whereby Sabinet Online would in future provide services to SABINET and its members. SABINET, together with some of its individual members, has a controlling shareholding in Sabinet Online and still own the SACat. The objectives of

SABINET are continuing through Sabinet Online. Many tertiary institutions in South Africa became shareholders in Sabinet Online.

Sabinet Online functions in a business environment where the Internet and the World Wide Web have become the standard mode in the delivery of information. There is greater focus on product development, client support and training, and marketing. The management philosophy is to

1. Develop products and services that will ensure optimal satisfaction of clients' needs;
2. Provide shareholders with an acceptable return on investment;
3. Offer its staff opportunities for personal growth and development; and
4. Make a significant contribution to developing and raising the level of the South African community at large.

## 6 The Dawn of a New Era

The formation of regional library consortia and their receipt of funds from The Andrew W. Mellon Foundation for new technologically advanced library systems have placed unprecedented demands on Sabinet Online since 1997. The libraries in these consortia, having been upgraded to a more advanced technology, found themselves outgrowing the limited functionality offered by the existing SACat infrastructure and functionality.

At the time, the SACat urgently needed to be upgraded, as it was functionally and technologically outdated. In fact, it was so outdated that certain consortia were unable to use it or were not prepared to pay for the use of such an outdated service. This situation has been exacerbated by the changing needs of users who required more sophisticated solutions.

The five library consortia have, to a greater or lesser degree, discussed plans for a regional union database internally and with Sabinet Online, since any decision taken by Sabinet Online on a national solution would affect the decisions of the consortia. Preliminary discussions were held with The Andrew W. Mellon Foundation, which encouraged Sabinet Online to seek a nationally acceptable solution.

During 1997, Sabinet Online started to work on a strategy for building a national information infrastructure, which will not only complement and

interface with the various library systems of the library consortia, but will also serve the needs of the wider library community throughout Southern Africa who are not members of these consortia.

The strategy was based on the original purpose of the establishment of SABINET in 1983, which was to establish and support a national resource sharing infrastructure, fully integrated with local and regional infrastructures, by means of

1. A national union catalog of South African bibliographic records and holdings of high quality that will support shared cataloging and acquisitions and eliminate duplication of effort and costs; and
2. A national interlending and circulation system that will facilitate mediated and unmediated transactions on a local, regional and national level.

It was evident that Sabinet Online had a unique role to play in combining all library initiatives in South Africa into an integrated national information infrastructure and in ensuring a high level of resource sharing in the country. Although the country does not have enough role-players or enough combined resources for the establishment of independent regional catalogs, it was clear that the temptation to do so was always there, which could have led to the alienation of the regions from one another. It was evident that South African libraries needed to cooperate even more closely than before, since their ability to purchase new material had been severely curtailed by budget cuts, high price increases and the poor exchange rate of the Rand.

During 1998, strategies and models for cataloging and interlending were developed and discussed at regional users' meetings throughout the country, as well as separately with the library consortia. During discussions with the regional library consortia, it became evident that there was a considerable overlap in the requirements for regional union catalogs and the SACat initiative. For example, GAELIC (GAUTENG and Environs Library Consortia) was urgently seeking a software solution for its resource sharing and shared cataloging needs, but was aware that it would have great difficulty in paying for both its own regional union database as well as online access to the new national union database. There was therefore an urgent need to avoid unnecessary duplication and costs, and to optimize existing and possible future funding.

Urgent discussions were needed between Sabinet Online, the regional consortia and representatives of other key library sectors such as the national libraries and the public/provincial libraries. Such a workshop was held on September 7, 1998, with the objective of gauging the level of support for a national, rather than regional, union database. It was attended by all the regional library consortia, as well as the State Library and representatives from the Public Library sector. During this meeting, the strategic importance of a National Union Catalog for facilitating shared cataloging and inter-library loans was fully endorsed, and Sabinet Online was assigned the task of obtaining funds for the establishment of a redesigned national infrastructure and SACat.

The detailed requirements were compiled with the assistance of some consortium members, and were widely distributed to all users for comment. This was done in an effort to involve both consortium and non-consortium users. Throughout the process, it became clear that the consultation of all parties involved and efforts to ensure their commitment were of the utmost importance.

These efforts resulted in the presentation of proposals to the Foundation<sup>8</sup> on October 10, 1998 and November 3, 1998 to support a strategy for national resource sharing in Southern Africa. This resulted in a two-phase project, which was initiated during 1999.

## Phase 1

In the first phase, it was decided that

1. The current cataloging procedures be replaced with the OCLC Prism service, which allowed users to do original cataloging, upgrade records, download high quality bibliographic and authority records for copy/shared cataloging. These various types of records would also be downloaded and housed on the national union database;

---

<sup>8</sup> P. F. Malan and D. L. Man, SABINET Proposal to The Andrew W. Mellon Foundation, October 1998.

2. The SACat on ERUDITE (in SAMARC) be replaced with a National Union Database of bibliographic records and holdings in USMARC and housed on a technologically advanced library system. The recommended system was the INNOPAC system by Innovative Interfaces, Inc.;
3. The new National Union Database was to contain bibliographic records of a high standard to facilitate shared cataloging through electronic data interchange. It was hoped that it would enable libraries, particularly the consortia, to eliminate duplication of cataloging, to become more efficient, and to cut costs. All original cataloging was to be done on the OCLC Prism system, and a copy of these records, as well as those copied from OCLC, would be housed on the local INNOPAC system. It was further decided that this combined and integrated service be called SabiCat.
4. It was further suggested that authority file upgrading be done on the existing database by external experts before loading the data into the new SACat.
5. It was finally suggested that bibliographic records on the old SACat be matched against the Worldcat database and upgraded to a higher quality, and that as many duplicate records as possible be removed.

## Phase 2

The plan was to replace the current interlending system on the SACat with a technologically advanced interlending system. As envisaged, the system was to make provision for requesting, supplying, administrative, statistical, and financial functions for returnable items, as well as photocopies. The interlending system was to be fully integrated into the SACat database, which would be housed on the INNOPAC system to provide the cataloging model implemented during Phase 1. After much investigation and several consultation sessions with the interlending community in South Africa, it was decided that the DRSS (Distributed Resource Sharing Software) from OCLC be implemented. The software, which could be accessed via a Web-based interface, was based on the functionality of the current OCLC ILL system. Although the specific software was only running in certain test phases in certain US consortia, Sabinet Online was comfortable with the

decision to implement the software, on the basis of its positive experiences with OCLC concerning the latter's ability to deliver on its promises.

### Implementation of Phase 1

The implementation of this phase of the project consisted of many aspects, ranging from hardware implementation to migration of the user community onto the new platform. The project was to be handled by a group of five staff members from Sabinet Online and the various staff members from the vendor organizations. The project was further conducted according to a project plan with certain deadlines, monitored on a regular basis by implementation meetings and followed up with feedback to users and vendors.

Although the failure of any aspect of a project of this magnitude can easily jeopardize the complete project, it became clear that certain aspects were more important than others. The following aspects were revealed as the most difficult:

1. The extraction of the SAMARC data from the old system, the conversion of the data to USMARC and the loading of the final upgraded bibliographic records into the newly implemented system; and
2. Training, which turned out to be a big problem; not only because of the many users needing to be trained within a large geographic area, but also because there were so many areas to be addressed during the training. The training consisted of teaching the users USMARC, teaching them how to use the new software in the form of the client software for connecting to INNOPAC and OCLC PRISM, and finally teaching them to adapt to a completely new workflow.

After successful implementation of the hardware and the configuration and implementation of the INNOPAC software, the process of loading the data, on which much work had been done up to that point, started in October 1999. According to data received from the vendor and calculations based

---

<sup>9</sup> P. F. Malan, *Progress on the Implementation Project of Phase-1 of the National Union Database of Southern Africa*, November 1999: 5.

on loading statistics from other similar implementations in South Africa, the Sabinet Online implementation team calculated that the loading should take no longer than three months. Since this period fell in the rather quiet November to January period, it was ideal for the project. Based on this timeframe, users were to be trained early in February to enable them to start using the new service early in the New Year.

By the end of January 2000, it became clear that the loading of the data had progressed much more slowly than anticipated. This was mainly due to the size of the catalog (3.5 million bibliographic records, over 8 million item records and 1 million authority records); however, problems also arose with the hardware, which resulted in further delay. The data loading was finally completed by May 2000, at which time good progress was already made by 19 classroom-style training sessions countrywide. The classroom-style training was followed up during June with 21 training sessions that took place onsite at user institutions. During these sessions, attention was not only given to the use of the services, but also to networking and related problems. This form of individual implementation proved very successful and resulted in more than 154 libraries adopting the service by October 2001. Today there are more than 170 libraries using the service, which are collectively downloading an average of 30,000 bibliographic records for shared cataloging purposes per month.

## Implementation of Phase 2

The second phase to install the ILL module began while the implementation of the first phase of the project was still in progress. Although it did not make much sense from a company and staffing point of view to have started with the project so early, there were not many options, since the old SACat on the ERUDITE platform was aging more by the day, with no further holdings updates taking place.

The project formally commenced during March 2000. This phase of the project was very different from the first phase, since over 400 libraries in South Africa participate in the ILL system, and because ILL is an interactive process among the various institutions, it was imperative for everybody to migrate to the new system at exactly the same time. The philosophy that was therefore adopted during the implementation was to

have the systems in place, to train all users in the shortest possible time, and then to set a date when everybody would start to process transactions on the new system, while the old system would be closed at that time.

Installation of the hardware and software was completed by May 2000. After the completion of installation, some users (mainly situated in the Gauteng region and who had received limited training) were given a three-week period to test the service by sending dummy requests to one another and to report any problems. It was later learned that not many users used this opportunity for testing, mainly because they were still very unfamiliar with the system and had other daily activities they were pressured to do. A very big problem, which surfaced at this early stage of implementation but was not taken into account during the planning of the project, was resistance to change. This problem would have been insignificant if the implementation team had placed more emphasis on involving libraries in the process of system choice and implementation and persuading them of the advantages of the project.

The training of the libraries began at the beginning of June 2000. During the three weeks that followed, six trainers conducted 37 training workshops of two days' duration. By the completion of the training phase, nearly 300 library staff members were trained in the seven biggest regions in South Africa.

On July 31, 2000, the ReQuest system went live and access was closed to the old ERUDITE system.<sup>10</sup> Although the changeover was irreversible, it did not happen without problems, and the months that followed were perhaps some of the most difficult experienced in many years, since almost the entire staff was either busy assisting users to adopt the cataloging service or to solve problems on the ReQuest system, which proved to require a lot of technical expertise to run effectively.

By September 2000, the ReQuest system was starting to become well established in the South African library market, despite the fact that many functionality problems and additional requirements surfaced at a time when training was still taking place on an ongoing basis. At the time, it was

---

<sup>10</sup> P. F. Malan, *Progress on the Implementation Project of Phase-1 and Phase-2 of the National Union Database Project of Southern Africa*, October 2000: 8.

decided to hold workshops on the new system with users who had already been trained to use it. The information obtained from these workshops was used to change/adapt the system to meet the users' specific requirements. The sessions and the opportunity for open discussion were welcomed by all libraries. These led to a list of requirements, which subsequently resulted in the complete redevelopment of the user interface and the incorporation of an IFM (Interlending Fee Management) system. This new user interface was finally implemented during March 2002, with much positive response from the library community.

Currently, the monthly average is 28,000 new requests on the system by more than 400 institutions with over 20,000 registered users.

## 7 Benefits and Cost Savings

Various studies and user experiences over the years have attested to the benefits and cost savings of shared cataloging. These benefits and savings are also fully experienced by South African libraries. The following are some benefits relevant to the South African library environment:

1. Shared cataloging, as compared to original cataloging, enhances the timeliness and productivity of technical services within the library. This not only means that books purchased are added to the online library catalog and available for circulation much faster, but it also contributes to costs savings;
2. By making use of the shared cataloging facility, library staff have access to an increased number of cataloging records, again contributing to the savings derived from not having to upgrade many records downloaded from the central shared cataloging service;
3. The use of the shared cataloging service has made library cooperation possible among libraries in the South African region. Through the availability of WorldCat, libraries now also share resources with the international library community and are part of the international cataloging fraternity; and
4. The cooperation in shared cataloging by the library community is contributing to the constant updating of library holdings on SACat,

which in turn is enhancing the sharing of resources through inter-library loans in the region. Resources are also now available online much more quickly for ILL because of shared cataloging.

The use of the service increased rapidly after the first year of implementation. This clearly indicates the value that libraries are deriving from the service. Factors such as the implementation of better networking infrastructures and local library systems have further contributed to the increase in the use of the service.

The following statistics were recorded for the period January to November 2001 (Table 1).

**Table 1. Basic Statistics**

	<b>WorldCat</b>	<b>SACat</b>	<b>Total</b>
<b>Records downloaded</b>	91,674	235,725	327,399
<b>Average searches</b>	347,614	895,755	1,243,369
<b>New records created</b>	10,297		10,297

The above usage statistics indicate that 327,399 records were downloaded to local library systems during the reported period. This represents considerable savings because the only alternative would have been to create these records at the institutional level from scratch.

Since a vast number of bibliographic records were downloaded from SabiCat and therefore not cataloged by libraries, cost savings for the country must have been considerable. To determine the cost savings, see studies done in the US.<sup>11</sup> These studies conclude that the average cost,

---

<sup>11</sup> Dilys E. Morris, Colin B. Hobert, Lori Osmus and Gregory Wool, *Library Resources and Technical Services, Cataloging Staff Cost Revisited*, Vol. 44 (April 2000): 74–76; J. Matthews, *Information Outlook, The Value of Information in Library Catalogs* (July 2000): 5.

including systems, administration and staffing, of original cataloging of a monograph is \$44.81 (R537.72), while the cost of copy cataloging of a similar item is \$12.22 (R146.64). Comparing the cost of original cataloging against copy cataloging (using the 327,399 records downloaded through copy cataloging as in our sample), we should obtain a fair estimate of cost savings for the country over the recorded period:

**Table 2. Cost Savings**

	<b>Cost per record (R)</b>	<b>Total records</b>	<b>Total cost (R)</b>
<b>Original cataloging</b>	537.72	327,399	176,048,990
<b>Copy cataloging</b>	146.64	327,399	48,000,789
<b>National saving</b>			128,048,201

The above calculation indicates a saving of R 128,048,201 to the library community as a result of shared cataloging instead of original cataloging. Even if costs in South Africa (e.g. salaries) are only one-third of those in the US, where the original studies were undertaken, the savings would still amount to R 42,682,733 (R 128,048, 201/3).<sup>12</sup>

## 8 Conclusion

Although some difficulties were encountered, the two phases of the project were completed in about a two-year period. Enhancements to the SACat and supported services will continue, and so will the training of additional users. Very valuable lessons were learned not only during the nearly twenty years of Sabinet Online's existence, but also during the implementation of the new systems and services. Some

---

<sup>12</sup> Morris, Hobert et al., and Matthews.

of the most valuable observations and lessons in the South African context were:

1. Standards are very important, and adherence to standards contributes to better and more effective resource sharing. The development of the SAMARC standard, although seeming to be a good decision while South Africa was in isolation, proved to be the wrong decision, since it limited the sharing of resources with the international library community and prevented the adoption of international technologies.
2. The development of a unique SAMARC standards-based union catalog system for the South African library seemed a very good decision during the mid-1980s; however, it failed due to poor planning and project management. Nevertheless, the failure of this project contributed to the long-term survival of shared networking in South Africa. If this project had succeeded, libraries in South Africa would have been left with an apparently perfect solution. However, this solution would ultimately have been unaffordable, due to the outdated mainframe technology on which it was based.
3. The implementation of new technology platforms and international standards brought about a new era of cooperation and resource sharing among South African libraries that had never been known before in the industry. This can largely be attributed to the technology that enabled these processes and libraries, through the formation of regional consortia that organized and forced their members to cooperate more effectively.

It is clear from the usage of the service and from some simple cost comparisons based on the use of shared cataloging that the service is of tremendous value to the South African library community. Without external funding, the library community in South Africa would not have been able to enter this new era of computerization and collaboration, a position which would have had an unfavorable impact on its long-term survival.

# **Chapter 20**

## **Regional vs. National Union Database Development: The GAELIC Perspective**

D. L. Man and Lettie Erasmus

### **1 Union Database History in South Africa**

South Africa has a long history of union catalog development for interlending and resource sharing purposes, beginning in 1912 with the compilation by A. C. G. Lloyd, Chief Librarian of the South African Public Library, of the Catalog of Serial Publications Possessed by the Geological Commission of the Cape Colony, the Royal Observatory, South African Association for the Advancement of Science, South African Museum and South African Public Library. The list consisted of some 1,030 periodical titles with a scientific bias, and was gradually supplemented in later revisions by the holdings of further libraries and scientific institutions throughout South Africa. By 1927, the number of contributing libraries had increased to 44 and the number of titles to 3,117.

It was clear by this time that there was a need to include humanities periodicals, leading to the publication of the Catalog of Union Periodicals in two volumes, edited by Percy Freer. Volume 1, Science and Technology, was issued in 1943 and again in 1949 and 1953, while Volume 2, The Humanities, was published in 1952. Two aims of this publication were to encourage libraries to amalgamate fragmentary holdings, and to eliminate unnecessary duplication. This was followed by a printed union list of periodicals in the whole of South Africa for the

period 1943–1952 which continued as Periodicals in South African Libraries (PISAL) from 1961–1972.<sup>1</sup>

The change of format to microfiche made it possible to produce a union list of monograph holdings. The State Library began producing the South African UNICAT in June 1972, which was a list of all monograph acquisitions of Southern African libraries with International Standard Book Numbers (ISBN). This was supplemented in 1975 with the publication of the Joint Catalog of Monographs 1941–1971, consisting of 2,139 microfiches in four boxes. These two sets of microfiches were combined in 1978 as the South African Joint Catalog of Monographs 1971– and published quarterly in author, title and UNICAT sequence. An interesting fact is that with the publication of the UNICAT in 1972, South Africa became the first country in the world to have a national union catalog of monographs based on ISBN and appearing on microfiche.

The periodicals catalog PISAL was also converted to microfiche format from 1974 onwards and published annually. Simultaneously with all these union catalogs, many subject and national bibliographies were being produced, making South Africa the most thoroughly documented African country south of the Sahara.<sup>2</sup>

The number of microfiches that had to be produced was cumbersome, and made it imperative that these catalogs be automated. Discussions concerning the establishment of a computerized national union catalog began in 1979 under the leadership of the National Library Advisory Council and the MARC Working Group. The use of SAMARC as an input and communications format in the South African Bibliographical and Information Network (SABINET) was seen as a prerequisite for coordinated computerized resource sharing.<sup>3</sup>

---

<sup>1</sup> Elizabeth Hartmann, in conjunction with J.I.Snyman, *The History of PISAL and its Forerunners* (Pretoria: CSIR, 1980) (CSIR Special Report CSTI 28).

<sup>2</sup> Reuben Musiker, *South African Bibliography*, 2<sup>nd</sup> ed. (Cape Town: David Philip, 1980).

<sup>3</sup> *SAMARC: South African National Format for the Exchange of Machine-Readable Bibliographic Description*, ed. Ilse van Niekerk (Pretoria: Working Group for Bibliographic

The founding of SABINET in 1983 and the development of the South African Union Catalog in SAMARC format are discussed in detail by Pierre Malan.<sup>4</sup> The many problems encountered during the creation of this national union catalog not only affected developments at SABINET, but also developments at user libraries.

## 2 Union Database Expectations and Disappointments in the 1980s and 1990s

There were high expectations among libraries, particularly academic libraries, for the development of a national union database. Not only would inter-library loans and resource sharing be made easier through the elimination of hundreds of out-of-date microfiches, but libraries could catalog centrally and download records into their own (often in-house) systems. Initial participating libraries had to sign an agreement to support SABINET for 10 years, and many of these libraries did so willingly.

The new union database was also seen as a way of keeping up with library automation in the rest of the world and reversing the impact of sanctions, because apartheid had profoundly influenced the development of the higher education sector in South Africa. Before the changeover of government in 1994, higher education institutions were divided along racial, language and political lines. They did, however, cooperate in the form of inter-library loans, as sanctions restricted the flow of information into South Africa. Sanctions also limited the choice of library systems available to libraries and to SABINET for developing the national union database, since system vendors were not able to do, or interested in doing, business in South Africa.

In 1985, SABINET embarked on an ambitious project to develop its own union database system. Libraries waited patiently for this new system, named Pythia, and in the interim used a system based on the Washington

---

Standards, Committee for a Computerised Cataloging Network [and] National Library Advisory Council, 1982). (National Library Advisory Council Report Series no. 24)

<sup>4</sup> P. Malan, "A National Union Catalogue for Shared Cataloging and Resource Sharing for Southern African Libraries," this volume.

Library Network (WLN) software. It was expected that Pythia would be available within three years, but as time went by there was deafening silence about progress on the part of SABINET. Prototype screens for the searching function were only made public around 1989, and were cumbersome and difficult to use. SABINET users were not happy, and suspected the presence of deeper problems when consultants were sent to some libraries to solicit their views on Pythia's response times, usability, etc. It came as little surprise when the development of Pythia was stopped in late 1990.

After this eight-year wait, individual libraries and SABINET had to start all over again. Academic libraries were forced to buy new library systems or keep developing their own in-house systems. Because of sanctions, the choice of available systems was limited to locally developed, affordable ones. Instead of centralized cataloging through a national union database, libraries went their own way and continued with original cataloging. It is difficult to speculate with hindsight how far South Africa's national union database development would have developed by now if Pythia had been successful, or if there had been no sanctions to limit SABINET's choice to a locally developed system.

### 3 Background to the Development of GAELIC

The Gauteng and Environs Library Consortium (GAELIC) was formed in April 1996 under the umbrella of its parent body, the Foundation of Tertiary Institutions of the Northern Metropolis (FOTIM). By that time, South Africa was post-apartheid and sanctions had been lifted. Technology had also changed substantially, and a number of academic libraries were investigating the possibility of purchasing new library systems locally, or overseas if they could afford them. The offer by The Andrew W. Mellon Foundation to fund common library systems within legally constituted academic library consortia was seen as a golden opportunity to leapfrog to technologically advanced library systems. Five library consortia availed themselves of this opportunity, namely GAELIC, CALICO, FRELICO, SEALS and eSAL.

GAELIC consists of 16 academic libraries (ten universities and six technikons) from the three northern provinces of South Africa, namely Gauteng Province, Limpopo Province and North West Province. These institutions, which had little contact during the apartheid years except at the

university librarian level and through inter-library loans, were now prepared to put aside political, racial and language differences and work together to share resources and staff expertise, as well as to reap the benefits of a common library system. There was great disparity among the institutions in terms of size, resources, and expertise, leading to the terms Historically Advantaged Institutions (HAIs) and Historically Disadvantaged Institutions (HDIs). The HDIs were Black institutions set up by the apartheid regime and were mostly situated in outlying regions, and the creation of a consortium provided the opportunity to lessen these disparities and to extend the collaboration among the members.

An important issue for the implementation of a common library system was the system architecture to be used within GAELIC. The architecture chosen could influence the choice of a library system, and vice versa. Factors to be taken into account included the size of GAELIC, the autonomy of the institutions, the lack of network stability, and the high cost of Internet connectivity on and between campuses.<sup>5</sup> Discussions revolved around several models, because each model had cost implications in terms of the size and quantity of servers and the number of software licenses. The following were some of the models discussed.

#### 4 System Choice

After extensive negotiations, the vendor of choice offered favorable pricing for Model 2 to be implemented. These separate systems allowed for a faster rate of implementation, since less consensus was needed for system setups. No databases needed to be merged nor duplicates sorted out, thus leaving institutions free to implement on their own when they were ready.<sup>6</sup>

---

<sup>5</sup> H. M. Edwards, "South Africa's GAELIC: the Gauteng and Environs Library Consortium", *Information Technology and Libraries*, 18 (3): 123–129. Special Issue: *Library Consortia around the World*, guest ed. John F. Helmer.

<sup>6</sup> D. L. Man and L. Erasmus, "Implementing a Library System in a Consortium: the GAELIC Experience", *Proceedings of the Conference on the Provision of Information in Southern Africa*, University of Pretoria, 20–21 August 1998: 134–136.

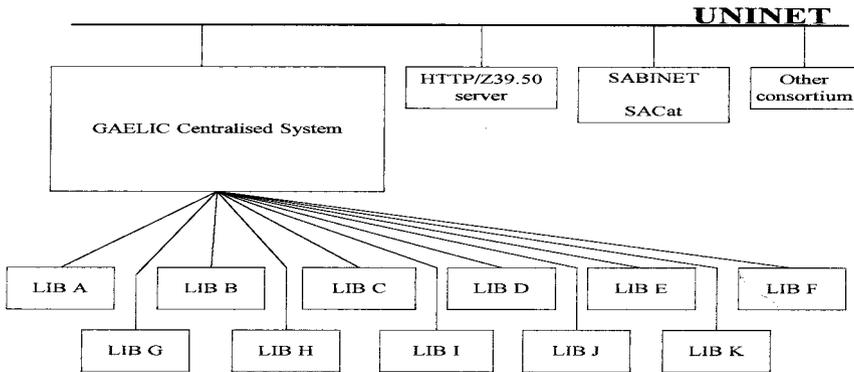


Figure 1. Model 1: Centralized System\*

\* One copy of the software package is loaded onto one large central server, to which all separate libraries are linked via Uninet (now called TENET). This lowers software costs, but requires a strong and robust network with high bandwidth and adequate redundancy. Hardware platform costs are high because of the very large scalable server required. This model would automatically result in a union database.

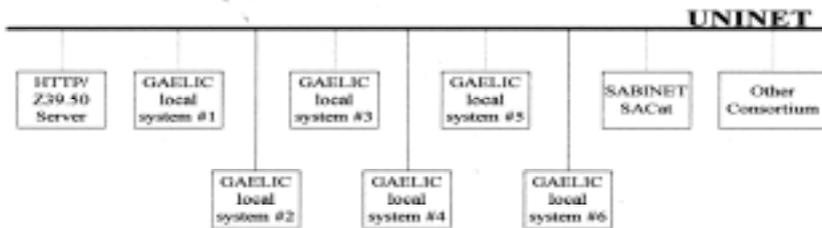


Figure 2. Model 2: Distributed System\*

\* Each of the 16 libraries would operate completely independently on its own separate server with its own version of the software package. High software and server costs, but less dependency on network stability and sufficient bandwidth. No union database is provided for.

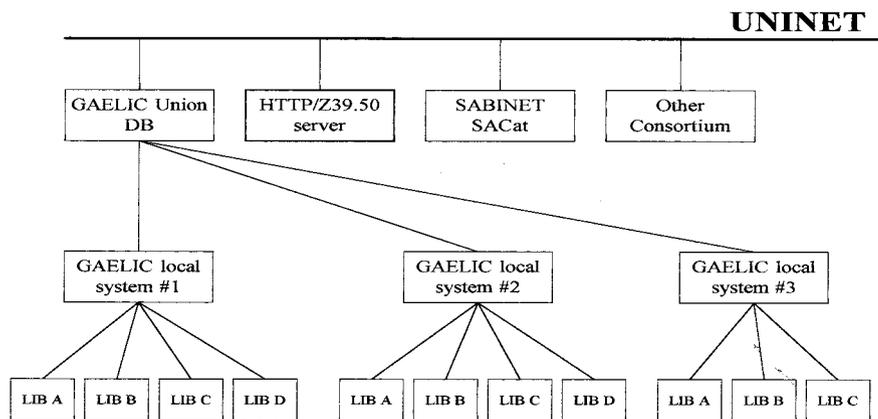


Figure 3. Model 3: Regionally Distributed Clusters with a Union Database\*

\* A compromise between Model 1 and Model 2, as some servers and software are shared, thereby reducing the cost of the software package and hardware platforms. Additional cost for union database.

The INNOPAC library system<sup>7</sup> was chosen, and implementation began in mid-1997 with Phase 1, comprising six libraries. Phase 2 followed in mid-1998 with eight libraries, including two libraries from the adjacent consortium FRELICO, the Free State Library Cooperative. Phase 3 began in mid-1999 with four libraries in outlying areas. Members of the phases (and the estimated number of titles in their collections at time of implementation) were as follows (Tables 1–3):

<sup>7</sup> The INNOPAC system is now called Millennium, and is developed by Innovative Interfaces Inc., Emeryville, California, USA.

**Table 1. Phase 1 Libraries**

<b>Institution</b>	<b>Titles</b>
Technikon Northern Gauteng	31,000
Technikon Pretoria	70,000
Technikon South Africa	50,000
Technikon Witwatersrand	50,000
University of South Africa (UNISA)	980,000
University of Witwatersrand	500,000

**Table 2. Phase 2 Libraries**

<b>Institution</b>	<b>Titles</b>
Medical University of South Africa (MEDUNSA)	33,000
Potchefstroom University for CHE	350,000
Rand Afrikaans University	338,000
University of Pretoria	356,000
Vista University	85,000
Vaal Triangle Technikon	38,000
Technikon Free State (FRELICO)	N/A
University of Free State (FRELICO)	350,000

**Table 3. Phase 3 Libraries**

Institution	Titles
Technikon North West	12,000
University of North West	80,000
University of the North	200,000
University of Venda	46,000

The speed of implementation was a considerable achievement, but GAELIC was still left with the problem of the union database.

## 5 The Need for a GAELIC Union Database

As could be seen from the models discussed in the preceding section, the incorporation of a union database had been part of GAELIC's planning from the outset, since it fulfilled its vision of creating a virtual library with local service interfaces, forming part of a global information community, for clients in Gauteng and its environs.<sup>8</sup> The union database would facilitate resource sharing and shared cataloging among the 16 members. An early decision by GAELIC was to provide free inter-library loans among its members, the justification being the need to assist smaller libraries, to render a cost-effective service and to ensure the free flow of information to researchers and students for the benefit of the consortium and the country as a whole. It was recognized that there would be net lenders who supplied more documents than they received, and that transactions should be monitored so that imbalances could be addressed.<sup>9</sup>

---

<sup>8</sup> Gauteng and Environs Library Consortium. <http://www.gaelic.ac.za/backg.html>.

<sup>9</sup> H. Visser, "The Free of Charge Document Supply Agreement within the GAELIC Consortium," Paper presented at the 7th International Interlending and Document Supply Conference, Ljubljana, Slovenia, October 2001.

It was hoped that a union database would reverse the amount of effort put into duplicate original cataloging by the institutions, which did 80% original cataloging and 20% copy cataloging in Phase 1 and Phase 2 libraries. Before a union database could be implemented, however, there were still many issues to be resolved within GAELIC.

## 6 GAELIC's Options for a Regional Union Database

With the completion of implementation of the INNOPAC library system in the six Phase 1 libraries in 1998, the need to amalgamate their bibliographic records and holdings together in one centralized database became urgent. Discussions revolved around the availability and choice of software, cost and sustainability, specifications, and integration with the databases of other consortia in addition to the SACat. INNOPAC offered two products for union databases:

1. The INN-Reach system, developed originally for OhioLink, which was excellent for resource sharing, inter-library loan transactions and automatic upgrading of bibliographic records on local systems, but did not allow for centralized cataloging because this was done on OCLC WorldCat. The consortium would need to purchase the INN-Reach software as well as a special module for loading onto local INNOPAC systems to enable them to integrate fully; and
2. The new software being developed for the National Library of Taiwan ("Taiwan version"), which would allow for centralized cataloging but was not yet ready. This system could integrate with all library systems that accommodated electronic data interchange.

At the same time, Sabinet Online, the newly formed for-profit arm of SABINET, was seeking a replacement for the SACat, which was by then technologically and functionally out of date and unable to keep up with the advanced systems being implemented by the consortia. Furthermore, the SACat was costly to maintain and had poor-quality records with many duplicates as a result of the lack of authority files and lack of quality control, as well as the decision to choose completeness of holdings over quality of bibliographic records.

The symbiotic relationship between SABINET and the consortia was recognized by the Mellon Foundation, which felt that it was important that a workshop be convened between these parties to gauge the level of support for a national, rather than regionally based, union database. A unified approach would realize the Mellon Foundation's objectives for national library collaboration in South Africa.<sup>10</sup> At that time, it was estimated that GAELIC as a regional consortium held 40% of South Africa's information resources, while CALICO held 30%.

At a joint workshop held on September 7, 1998, the following requirements for a national infrastructure to support shared cataloging and interlending were defined:

- The need for a shared, cost-effective document delivery system in South Africa;
- The importance of an affordable national information system;
- Less original cataloging. Shared cataloging on a high-quality, cost-effective system should be encouraged;
- The functionality of the system is more important than the platform on which it is housed;
- There should be end-user benefits and end-user access; and
- South Africa should have a joint collection development strategy based on a distributed national collection.

As a result of this workshop, the Mellon Foundation supported a proposal for funding for the redevelopment of the SACat and an interlending system, both of which currently resided on an ERUDITE system.

GAELIC and Sabinet Online were both faced with the problem of not being able to find union database software that was able to fulfill the requirements for both resource sharing and shared cataloging. Various options were investigated, and Sabinet Online finally decided to choose INNOPAC's 'Taiwan version' for shared cataloging and OCLC's Document Retrieval and Supply System (DRSS), locally named ReQuest, for interlending. It was also decided to use OCLC WorldCat as a cataloging utility for original and copy

---

<sup>10</sup> GAELIC National Union Database. See [http://www.gaelic.ac.za/national\\_union\\_database.html](http://www.gaelic.ac.za/national_union_database.html).

cataloging, and that no original cataloging would be done on the new SACat in order to maintain it as a database of high quality. GAELIC dropped its search for union database software and adopted the Sabinet Online strategy outlined above. As GAELIC could not afford its own union database, it had to utilize the scoping functionality on SACat as well as MagNET, SABINET's search and retrieval system for end-users, to limit searches to GAELIC holdings only.

While these decisions were being made and the new SACat was being implemented in 1999, GAELIC's phased implementation was continuing unabated, and institutions were getting used to doing original cataloging on their local INNOPAC systems—and probably getting back to their old bad habits. They also started using other options, eg the Z39.50 option in the cataloging module, to the detriment of shared cataloging and complete holdings on SACat.

## 7 The GAELIC Scope on SACat: Pros and Cons

The scoping product in INNOPAC allows users to confine their searches at the outset to a subset of the database, such as location. Sabinet Online decided to have five location scopes on SACat:

- GAELIC/FRELICO together;
- CALICO ;
- SEALS;
- ESAL;
- South African National Bibliography (SANB).

There were many benefits to GAELIC for using the SACat as its union database, including:

- Not having to pay for its own software, thereby removing the problems of affordability and sustainability;
- Automatically being part of a national collaborative effort for shared cataloging and resource sharing;
- The ability to limit searches to GAELIC libraries only; and
- The ability to identify GAELIC holdings for inter-library loan purposes.

However, there were also a number of disadvantages in giving up the idea of a GAELIC union database, including:

- Lack of control over the union database, because it is administered solely by Sabinet Online. GAELIC has no access to its own headings reports, cannot draw consortium-level or institutional-level statistics, cannot create its own lists for error detection or checking of holdings, etc. A key question is whether Sabinet Online will be able to render a statistical and reporting service, and at what cost;
- Lack of control over the quality of bibliographic records in the GAELIC scope, particularly because this could affect the success rate of searches. GAELIC libraries would like to do their own quality control within the GAELIC scope;
- End-user access to SACat through MagNET in order to see the GAELIC scope will have cost implications for libraries, since not all libraries allow their users access to MagNET;
- Inter-library loans are done on the ReQuest system via MagNET, yet each GAELIC library has an inter-library loan module for pre-requests by end-users. To make full use of the latter, some libraries would like to have an interface built between the two systems;
- End-user access to electronic resources within a union database is a complex issue. Not only does each institution have its own URL and/or IP restrictions, but each consortium or institution has its own access agreements with the vendors. The SACat still needs to address these issues;
- The Z39.50 option on MagNET is not properly utilized because of firewall restrictions and/or network problems at the various GAELIC sites;
- Other Z39.50 problems are the large number of duplicates retrieved in a search of several million records, and that 'See References' in the authority record are not taken into account; and
- The holdings format is not adequate for collection development and statistics. The holdings statement needs more detail, especially serials holdings for inter-institutional rationalization. At present the holdings

record on SACat consists of the following string of subfield codes: \$aCall number \$bLocation \$cVolume, etc. \$xLoan restriction.

These subfields cannot be used for statistical reports for collection development purposes. It would be desirable to have more subfields to accommodate media/format, code for subjects, code for identifying publisher, code for identifying titles that form part of existing/future consortium agreements, code for identifying and retaining the last holding of a serial title in the region or South Africa, etc.

## 8 GAELIC's Cataloging Problems on the SACat

In terms of functionality, the new SACat had a number of problems that had to be addressed with the vendor, including the cumbersome way of updating holdings statements.

Ironically, instead of reaping the rewards of using the same software, GAELIC libraries have more problems with copy cataloging than non-INNOPAC libraries. This is mainly because of the functionality problems of the dual connection that INNOPAC libraries use for downloading records from the SACat into their local systems. When the dual connection is open, duplicates are generated, since there is no matching on ISBN or OCLC number. The item information must then be transferred from the incomplete record to the full record and the duplicate deleted.

There are three possible procedures for cataloging:

Method 1:

1. Open Dual Connection on Cataloging Workstation.
2. Open local and central connection.
3. Search central database, SACat, for bibliographic / authority record.
4. Find bibliographic / authority record.
5. Save to local database.
6. Display screen to set institution holdings for bibliographic record on SACat.
7. Search local database, add local item data and other relevant data to bibliographic record and save.

#### Method 2:

1. Find no bibliographic/authority record on SACat.
2. Open OCLC session via CatME to WorldCat.
3. Search WorldCat for bibliographic/authority record.
4. Find bibliographic/authority record and export to SACat (the system automatically validates the record and sets the holdings) via a networked interface.
5. Repeat steps 1 to 7 in order to make the bibliographic records available on the local database.

#### Method 3:

1. Find no bibliographic/authority record on SACat/WorldCat, or records need to be upgraded.
2. New full bibliographic records are created on OCLC WorldCat.
3. New authority records are created only by NACO participants on OCLC WorldCat.
4. Bibliographic or authority records are upgraded on OCLC WorldCat.
5. Export new / upgraded bibliographic / authority records to SACat via a networked interface.
6. Repeat steps 1 to 7 in order to make the bibliographic / authority records available on the local system.

## 9 GAELIC's Cataloging Decisions for National and International Compatibility

GAELIC's aim of promoting shared cataloging could only become reality if common standards and practices were established, and the GAELIC Cataloging and Technical Services Workgroup (GCATS) had a lot of work to do in this area. A survey carried out among the first twelve members of GAELIC in 1996 revealed a great diversity of cataloging practices:

- Number of library systems: 5 (5 ERUDITE, 3 Stylis, 2 ITS, 2 in-house);
- MARC system used: 10 SAMARC, 1 USMARC, 1 UKMARC;
- Language of catalog: 10 English, 2 Afrikaans;

- Format of authority records: 1 USMARC, 11 local system format;
- Form of names: 4 use only initials and surname, others use full names;
- Contribution to SACat: 6 contribute bibliographic records and holdings, 2 contribute holdings only, 2 contribute incomplete holdings, 2 do not contribute;
- Extent of original cataloging: 9 perform all original cataloging, 2 download from SACat only, 1 downloads from OCLC WorldCat and SACat.

These differences needed to be resolved before there could be any talk of shared cataloging or the merging of catalogs in a union database. In choosing the library system, the functionality rather than the particular MARC system was the overriding consideration, namely that it should be Web-based, have the latest bibliographic and technical developments, quality control mechanisms and electronic data interchange, and incorporate international standards such as ANSI, NISO, ISO and especially Z39.50. It was seen as a bonus that the INNOPAC system chosen was USMARC-based, as the SAMARC system was no longer being updated and did not include new technological requirements such as URLs for Internet linking or other requirements for new formats.<sup>11</sup> This decision by GAELIC to be the first consortium in South Africa to change to a USMARC-based system no doubt influenced other consortia and libraries to do likewise.

GAELIC members were keen to become part of the global library community, and GCATS took the following decisions to ensure that they conformed to international standards:

- Changeover from SAMARC to USMARC (now called MARC21),<sup>12</sup> since this would allow for greater use of copy cataloging
- Language of the library catalogs to be English

---

<sup>11</sup> D. L. Man and L. Erasmus, "Changing to Another MARC format: the GAELIC position," paper presented at Future MARC Format for South African Libraries, Pretoria, April 24–25, 1997.

<sup>12</sup> MARC Standards. MARC21. <http://www.loc.gov/marc/>.

- Cataloging rules and guidelines to be AACR2R with all revisions and updates,<sup>13</sup> Library of Congress Rule Interpretations,<sup>14</sup> OCLC bibliographic formats and standards<sup>15</sup>
- Library of Congress core records with a few local adjustments
- ALA-LC Romanization Tables: Transliteration Schemes for Non-Roman Scripts<sup>16</sup>
- MARC21 format for bibliographic records, authority records and holdings
- Library of Congress subject headings (LCSH) and Medical Subject Headings (MeSH). Where a local deviation is required, formal approval is sought from the Library of Congress, e.g. kwaito (music).
- Library of Congress Name Authorities on OCLC WorldCat.
- Classification systems are Dewey Decimal Classification (latest edition), Library of Congress Classification System and National Library of Medicine Classification System. Local deviations are not recommended except where there is a formal agreement to deviate because Dewey does not accommodate local needs satisfactorily, eg classical literature, African languages. GAELIC's proposed changes to the schedule for African languages have been incorporated in DDC Edition 21 (see Table 4).<sup>17</sup>

---

<sup>13</sup> *Anglo American Cataloging Rules*, second edition, revised 1988, under the direction of the Joint Steering Committee for Revision of AACR, eds. Michael Gorman and Paul W. Winkler (Ottawa, Canadian Library Association: ca. 1988).

<sup>14</sup> *Library of Congress Rule Interpretations* (Washington DC: Cataloging Distribution Service, Library of Congress, 1990).

<sup>15</sup> OCLC Bibliographic Formats and Standards. See <http://www.oclc.org/oclc/bib/about.htm>.

<sup>16</sup> *ALA-LC Romanization Tables: Transliteration Schemes for Non-Roman Scripts* (Washington, DC: Library of Congress, 1991). Approved by the Library of Congress and the American Library Association. Tables compiled and edited by Randall K. Barry.

<sup>17</sup> *Dewey Decimal Classification and Relative Index*, devised by Melvil Dewey, Edition 19, Vol. 2 (Albany: Forest Press, 1979); *Dewey Decimal Classification and Relative Index* devised by Melvil Dewey, Edition 21, Vol. 2 (Albany: Forest Press, 1996).

**Table 4. Dewey Decimal Classifications**

Dewey Decimal Classification Edition 19		Dewey Decimal Classification Edition 21	
African	-96	African	-96
Languages	-963 9	Languages	-963 9
Bantu languages	-963 9	Bantu languages	-963 977 5
Tswana	-963 9	Tswana	-963 978
Xhosa	-963 9	Xhosa	-963 986
Zulu		Zulu	

All these cataloging decisions have been debated with other libraries through the Sabinet Online Standards Committee and have been adopted as national standards for use in the SACat.<sup>18</sup> Having made these decisions, GCATS also had the task of implementing them. It arranged many training sessions in preparation for the implementation of INNOPAC in the various libraries, and also to raise the level of cataloging expertise in all the GAELIC libraries. Being the pioneer in MARC21 system implementation presented its own challenges, since the trainers had to train themselves before they could train others. These training sessions covered various areas, including MARC21 Bibliographic; MARC21 Authorities; cataloging of law publications, music publications, electronic publications and serials; assigning Library of Congress subject headings; using the INNOPAC cataloging module, and downloading bibliographic records from SACat and OCLC WorldCat. The standardized approach to cataloging policies and practices was seen as a key benefit to consortium membership and INNOPAC system implementation, in addition to training by experienced catalogers and sharing of ideas and expertise.

---

<sup>18</sup> SABINET Online Standards Committee.

See [http://www.sabinet.co.za/sabicatweb/sabicat\\_standards.html](http://www.sabinet.co.za/sabicatweb/sabicat_standards.html).

## 10 Authority Control and Participation in NACO

Most systems used by GAELIC libraries prior to 1997 did not allow for authority control, although some libraries used the Library of Congress name authorities and subject headings as guides. The SACat itself did not have an authority file, and could not be used as a reference source.

It was agreed that the proposal for donor funding should include database conversion from SAMARC to MARC21, and for the newly converted MARC21 database to be sent for authority headings matching and cleanup. The result of this process was that each of the libraries started off with a much cleaner catalog, as well as name and subject authority files for all the headings that matched an existing name or subject heading in the Library of Congress authority files. The hit rate for the GAELIC libraries varied between 67% and 87%. The more closely a library conformed to the Library of Congress's authority control practices, the higher the hit rate.

GAELIC also compiled the *Authority Control Manual and Policy Guidelines for GAELIC Libraries*<sup>19</sup> in 1998 to ensure the adoption of a standardized approach to authority control and to maintain the quality of the new authority files.

In 1999, GAELIC libraries decided to formally become participants in the Names Authority Cooperative (NACO), a part of the Program for Cooperative Cataloging (PCC) managed by the Library of Congress and consisting of nearly 400 cataloging organizations worldwide.<sup>20</sup> The reason for this was that GAELIC had agreed to accept the Library of Congress Name Authorities as the sole source of name authority headings, but this meant acceptance of the many incorrect headings for Southern African authors. Over the years, the Library of Congress had had little or no knowledge of Southern African languages or access to local reference sources, with the result that many South African headings were incorrectly established, e.g. N.P. van Wyk Louw appeared as Van Wyk

---

<sup>19</sup> *Authority Control Manual and Policy Guidelines for GAELIC Libraries*, prep. by Hester Marais, Ann van der Walt and Welna van Eeden (GAELIC, ca. 1998).

<sup>20</sup> Program for Cooperative Cataloging. NACO. See <http://www.loc.gov/catdir/pcc/naco.html>.

Louw, N.P. This has since been corrected by GAELIC to Louw, N.P van Wyk (Nicolaas Petrus van Wyk) 1906–1970. The first group of libraries received NACO training from a Library of Congress trainer in July 2000. Since then, GAELIC has been accepted as a member of PCC and acts as a funnel for interaction with Library of Congress.

**Table 5. NACO Statistics, October 1, 2000, to March 31, 2002**

	<b>Record Created</b>	<b>Records Changed</b>
<b>GAELIC</b>	2,551	934
<b>Rest of South Africa</b>	581	69

These NACO headings are posted on the GCATS listserv as well as the Sabinet Online Standards Committee website. GAELIC plans to expand its international activities by participating in NACO's series training.

## 11 GAELIC's Progress and Achievements

To measure GAELIC's progress from original cataloging to copy cataloging between 1997 and 2001, a subsequent survey was done and showed remarkable declines in original cataloging (see Table 6).

The survey conducted in 1997 among 11 GAELIC libraries in preparation for the implementation of the common library system highlighted the high percentage of original cataloging done by the GAELIC libraries on their local systems, although bibliographic utilities or services, e.g. SACat, Library of Congress MARC records, were available for copy cataloging purposes. Only four of the GAELIC Libraries explored the copy cataloging option, either by downloading bibliographic records from SACat (if their local systems supported the capability) or importing full Library of Congress MARC records from other sources.

**Table 6. Original Cataloging in 1997 and 2001**

<b>GAELIC libraries</b>	<b>% in 1997</b>	<b>% in 2001</b>
Medical University of South Africa (Medunsa)	10	20
Potchefstroom University for CHE	15	20
Rand Afrikaans University	100	20
Technikon Northern Gauteng	30	50
Technikon Pretoria	100	30
Technikon South Africa	100	10
Technikon Witwatersrand	100	20
University of Pretoria	100	20
University of South Africa (Unisa)	60	20
University of the Witwatersrand	100	20
Vista University	100	20
* Technikon North West	–	10
* Vaal Triangle Technikon	–	20
* University North West	–	20
* University of the North	–	100
* University of Venda	–	20

\* Became formal members of GAELIC in 1998 and were therefore not surveyed in 1997.

A follow-up survey conducted among the 16 GAELIC libraries at the end of 2001 showed that the percentage of original cataloging done by the GAELIC libraries changed dramatically. Thirteen of the GAELIC libraries reported that they were cataloging between 10% and 20% of the new items received originally. Only two libraries initially experienced no dramatic improvement

in their rate of original cataloging; however, that was mainly due to networking problems between their sites and the bibliographic utilities, which have fortunately been resolved. One library is experiencing problems and is doing no copy cataloging.

The implementation of the INNOPAC/Millennium system and the ease of downloading full bibliographic records from either SACat or OCLC WorldCat via the CatMe facility contributed to the lower rate of original cataloging done by the GAELIC libraries.

The general conclusion from the GAELIC libraries was that most of the original cataloging was required for non-English titles, South African published titles, local theses and dissertations, as well as very specific subject areas not covered in the bibliographic utilities, such as alternative health science.

**Table 7. Cataloging Statistics April 2001 to March 2002**

	New records added to WorldCat		Records copied from WorldCat to SACat	
<b>GAELIC</b>	3,459	35%	49,106	46%
<b>Rest of South Africa</b>	6,378	65%	58,451	54%

It is evident that GAELIC members have made good progress during the past five years in terms of shared cataloging, and there could be many more developments during the next five years as the GAELIC scope is developed. However, there is concern about the need for speed in loading GAELIC holdings onto the SACat for resource sharing purposes. One of the GAELIC member libraries has written a program for the batch loading of holdings onto the SACat, and this is being tested. The GAELIC scope will not be successful until all member holdings are loaded and kept up to date and accurate.

## 12 The Future

In June 2002, the long-awaited final report of the Department of Education was released, setting out a new model for higher education in South Africa that would

reduce the number of institutions from 34 to 21 through mergers and closures. Some institutions would remain as separate institutions.<sup>21</sup> It is proposed that GAELIC institutions will be reduced from 16 to 8, which may result in the merging of their databases. GAELIC's system architecture may therefore need to change to a new one, which will be a mixture of the earlier models discussed, i.e. separate systems as well as regionally distributed clusters.

In hindsight, GAELIC's decision to opt for separate servers and systems at each institution will actually help these new developments, because it will not be necessary to undo any databases, only merge already established ones. There will be a high initial cost in merging and deduplicating the databases and changing holdings statements, but in the long run there will be savings in maintenance costs. As far as the GAELIC scope is concerned, the current way of working with OCLC WorldCat and SACat will probably remain unchanged, but holdings statements will have to be updated. There are interesting times ahead, and GAELIC will have new challenges to face.

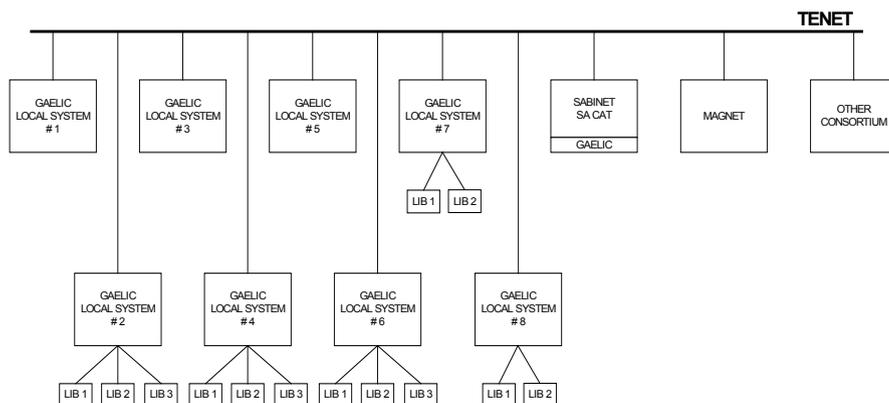


Figure 4. New GAELIC Architecture

<sup>21</sup> "Transformation and Restructuring: A New Institutional Landscape for Higher Education" (Ministry of Education) (June 2002).

See [http://education.pwv.gov.za/Policies%20and%20Reports/2002\\_Reports/he/transrestructuring/transformation\\_and\\_restructuring.htm](http://education.pwv.gov.za/Policies%20and%20Reports/2002_Reports/he/transrestructuring/transformation_and_restructuring.htm)

### 13 Evaluation

GAELIC's progress and achievements in terms of original versus copy cataloging and the pros and cons of giving up the idea of a regional union database in favor of a national union database are discussed in earlier sections. But how does the SACat measure up in terms of a real versus a virtual catalog, and the success or failure in meeting GAELIC's requirements for a union database and in cost savings? Have the right decisions been made?

SACat is a physical union database and meets the requirements of GAELIC librarians for cataloging and interlending purposes in the South African context. Although expensive to maintain and not current in terms of GAELIC holdings, the quality of the bibliographic records can be controlled. Access is stable and reliable, and there is no dependence on Z39.50 access, the problems of which were discussed above. End-users are able to access the SACat via MagNET and to request material through the linked ReQuest Interlending module. There is less dependence on high bandwidth, which in South Africa is expensive and not readily available.

In terms of cost savings, a record is copied from OCLC WorldCat once by a member library, and all other libraries make use of this same record. This saves staff time and OCLC costs. Libraries can then add their holdings symbols to this record on SACat, thereby allowing other libraries to borrow the item. According to a GAELIC survey, for the period 1997–2000, an average 35,300 documents were supplied annually within the consortium. Universities supplied 95% of these documents, and technikons 5%.<sup>22</sup> This is an indication of the extent to which the SACat and the ReQuest Interlending modules facilitate resource sharing within GAELIC alone, not taking into account the extensive resource sharing with libraries throughout South Africa. This level of activity will increase once the GAELIC libraries are able to load and update their holdings onto SACat and the GAELIC scope is fully utilized.

---

<sup>22</sup> Visser.

There is still a great deal of work that needs to be done on the GAELIC scope on SACat, and cataloging problems need to be sorted out, but it is important that the GAELIC libraries and Sabinet Online maintain the interest and goodwill to make the union database as accurate and up-to-date as possible.



## **Chapter 21**

# **Why the “Big Bang” Did Not Happen: The CALICO Experience**

Amanda Noble and Norma Read

As a result of preliminary talks held in April 1992 between the five tertiary institutions in the Western Cape, the Ford Foundation and the American Council of Education (ACE), a team of consultants visited South Africa later that year. The focus of the visit was to assess the level of interchange between the libraries of these five institutions, and to facilitate post-apartheid academic cooperative planning. It was hoped that these endeavors would gain financial backing from the Ford Foundation and others, with a view to greater support for teaching and research at the five institutions, thereby becoming a model for the rest of South Africa in encouraging other areas of cooperation within academe.<sup>1</sup>

The many years of isolation of South African scholars and information providers led to gaps in knowledge, information management and curriculum advancement, the redress of which is crucial for economic development. Current national emphasis is geared towards universal education at the primary level, adult literacy programs and information literacy, and this has resulted in cuts in state subsidies to tertiary institutions, together with an increase in the numbers of students, many of whom cannot afford escalating tuition fees. Dramatic rises in the costs of print subscriptions and electronic resources are further compounded by the drastic devaluation of

---

<sup>1</sup> Patricia Senn Breivik, Gary Pitkin and John Tyson, “The Western Cape Library Cooperative Project: Regional Planning for Post Apartheid Academic Development in the Western Cape,” unpublished report to the Ford Foundation, November 1992.

the rand—from R2 to £1 sterling in 1970, to R16 currently (October 2002), the decline in 2001 alone being 37%. Whereas First World countries experienced a 35% increase in the average cost of a social science journal between 1995 and 2000, the increase for academic libraries in South Africa was 203%. The imposition of VAT on ‘knowledge materials,’ together with the geographic isolation from the information centres of the world, means additional expenses on many items.

The five tertiary institutions in the Western Cape comprise two technikons and three universities, clustered within 50 km of each other. They have very diverse histories, but all of them are concerned with the transformation of higher education. The region was described in a recent report by the National Working Group on the Restructuring of the Higher Education System in South Africa as “one of the best-endowed provinces in South Africa as far as higher education is concerned,”<sup>2</sup> which further increases the obligation to provide major sustainable progress through the purposeful pursuit of strategic objectives. Research outputs, in the form of masters and doctoral graduates, and research publication units are amongst the highest in the country. In 1994, the two technikons generated more research funding from statutory councils and published more accredited articles than all the other South African technikons combined.<sup>3</sup> In 2000, the three universities in the region produced 28% of the research publication output and 25% of the doctoral graduate output of the public university plus technikon systems in South Africa.<sup>4</sup>

The Western Cape Tertiary Trust, which now operates as the Cape Higher Education Consortium, was formed in September 1993 to “facilitate and expand cooperation between the beneficiaries with regard to sharing of infrastructure, such as libraries, information technology, training of

---

<sup>2</sup> *The Restructuring of the Higher Education System in South Africa: Report of the National Working Group to the Minister of Education, December 2001*, Pretoria: Dept. of Education, 2002.

<sup>3</sup> J. A. Coetzee, “An Overview of the Western Cape Library Cooperative Project,” paper presented at the SAILIS Conference, September 21, 1994.

<sup>4</sup> *The Restructuring...*

personnel, as well as any other form of cooperation which may be beneficial to any of the parties.”<sup>5</sup> The Cape Library Cooperative (CALICO) is only one of the projects established by the Trust.

## 1 The Institutions

The Cape Technikon officially opened in 1923 but was only established by an Act of Parliament in 1979. It has faculties such as Built Environment and Design, Management, and Business Informatics, and offers more than 65 national diplomas and 45 BTech degree courses as varied as Horticulture, Librarianship, Hotel Management, Nature Conservation, and Parks Management—all of which have a heavy emphasis on practical components.

Peninsula Technikon's roots go back to 1962, when a steady growth in apprentices in a variety of trades led to the establishment of Peninsula Technical College. In 1972, the status of the institution was changed to the Peninsula College for Advanced Technical Education, and in 1979 this college became the Peninsula Technikon. The institution was granted partial autonomy in the early 1980s and full autonomy with the passing of the Technikon Act in 1993. Career-specific academic programs at the Peninsula Technikon are offered in three faculties: Engineering, Science, and Business. Short courses and opportunities for further education are offered through the Technikon's Centre for Continuing Education.<sup>7</sup>

A major change to the kind of education and training offered by the Technikon was brought about by the 1993 legislation, which expanded the qualifications offered to include degree courses (Bachelor of Technology, Masters in Technology and Doctorate in Technology). Each of the technikon has approximately 10,000 students.

---

<sup>5</sup> “Western Cape Tertiary Institutions Trust Memorandum of Agreement between the Trustees and the Councils,” September 14, 1993.

<sup>6</sup> See <http://www.ctech.ac.za/users/visitors.html>.

<sup>7</sup> See <http://www.pentech.ac.za/>.

Following the implementation of the University Act of 1918, the University of Stellenbosch was formed by the amalgamation of various colleges and schools, some of which dated back to the end of the 17th century. There are currently 17,000 students, 65% of whom have Afrikaans as their home language. It houses the Theological College for the Dutch Reformed Church, has faculties as varied as Military Science, Agriculture, and Forestry, and, as it is situated in the wine lands of the Western Cape, it has courses in Viticulture and Wine Biotechnology. Historically disadvantaged students comprise only one-third of the student body, but a diversity campaign with equity targets within defined timeframes was launched recently.<sup>8</sup>

The University of Cape Town, founded in 1829, is the oldest institution of the five, and in recent years was restructured into 6 faculties. Renowned as a 'liberal' university, the period 1960–1990 was marked by sustained opposition to apartheid. The student population is now 19,000, with a 48:52 black/white ratio and a rich diversity of students who come from some 70 different countries. Particular emphasis is placed on postgraduate studies, and 30% of students are enrolled in postgraduate programs. UCT is internationally recognized as one of Africa's leading research universities, currently having 14 top-rated scientists out of a national total of 45, a number of whom are recognized as world leaders in their field. An article that appeared in the *Financial Times* on May 11, 2002, included UCT in a list of 24 world universities chosen by university vice-chancellors as having an international reputation for excellence.<sup>9</sup>

The Extension of University Education Act of 1959 barred black people from attending institutions designated for white people, unless by special concession when alternative facilities were not available. In terms of its provisions, a number of colleges for specific designated categories were established the following year; thus, the University of the Western Cape began in 1960 as an ethnic college for 'colored' students. Following its establishment, the University College Western Cape was placed under the tutelage of the University of South Africa (UNISA) in Pretoria, and was run

---

<sup>8</sup> See <http://www.uct.ac.za/>, <http://www.sun.ac.za>.

<sup>9</sup> "A World Elite is Beginning to Take Shape," *Financial Times*, May 11, 2002.

by academics who supported racial separation and who saw their role as ‘white guardians’ of their ‘colored wards.’

UWC has worked hard to overcome its apartheid-driven origins, gaining full autonomy in 1984, and is committed to nurturing the cultural diversity of South Africa and to responding in critical and creative ways to the needs of a society in transition. Drawing on its proud experience in the liberation struggle, the university is aware of a distinctive academic role in helping to build an equitable and dynamic society. From an initial enrollment of 170, the student complement is now 10,000, drawn from all of the country's 11 language groups. UWC has grown from three to seven faculties, which comprise 68 departments and 16 institutes, schools and research centers, and was assessed recently by the Human Sciences Research Council as fifth out of 21 South African universities for humanities and social science research.<sup>10</sup>

A recent report of the Working Group on Higher Education has recommended that the number of universities and technikons in South Africa be rationalized from the current 36 to 21, and proposed that the Peninsula Technikon and the University of the Western Cape should merge to form one unitary ‘comprehensive institution’ offering both university-type and technikon-type programs. After some public debate, this suggestion has been rejected and a merger of the two technikons approved instead.<sup>11</sup> While recognizing what has already been achieved in the Western Cape, the draft report claims that “much more could be done with regard to the joint development and delivery of new academic programs, with regard to the coordination of existing programs to ensure the optimal use of resources and the satisfactory fulfillment of needs, and with regard to cooperation in the building of capacity where it is lacking or inadequate.”<sup>12</sup>

---

<sup>10</sup> See <http://www.uwc.ac.za/>.

<sup>11</sup> “A New Institutional Landscape for Higher Education in South Africa,” Report of the Minister of Education to the President, May 30, 2002.

<sup>12</sup> *The Restructuring...*

## 2 The Collections

The strengths of the library collections at UWC and the two technikons reflect the academic programs offered, and help to provide access to undergraduate texts for students who can ill afford to own their own copies. UWC has an audio-visual Self Access Learning Centre to assist students who come from disadvantaged educational backgrounds. UCT has the largest collection of the five, with a number of specialist branch libraries, and has unique research collections in areas such as Government Publications and African Studies. Its Rare Books Division houses what is thought to be the world's largest collection of fore-edge paintings. The recently established Knowledge Commons is the first in Africa, and provides undergraduates with a 'one-stop shop' for access to printed and electronic learning and research resources, plus office software to process their work.

As can be expected from its history, Stellenbosch has collections in Africana, Theology and Missiology. The Forestry Library has a unique collection of pamphlets covering all forestry, agroforestry, nature conservation and wood science disciplines. The Western Cape region is particularly strong in the performing arts, with UCT having schools of Drama, Ballet, and Music, both classical and jazz studies, while Stellenbosch has a Conservatory of Music. Together, the library collections that are reflected online amount to approximately 1.6 million bibliographic records.

## 3 Library Systems

The five institutions used four different library systems.

### BOOK Plus (Peninsula Technikon and University of Cape Town)

BOOK Plus, operated by Stowe Computing in Australia, could provide management information, financial data, borrowing statistics, ordering information, etc. At Peninsula Technikon, this system supported registration, circulation, cataloging, acquisitions, and serials functions, while at UCT the OPAC initially comprised only approximately 20% of the total collection

and was only used for searching. Circulation was via an Ontel system. Networking was not a component of the system, and downtime was considerable, with little or no local troubleshooting or support. A programmatic match of BOOK Plus and the Ontel system at UCT caused data corruption which necessitated manual unscrambling of items attached to incorrect bibliographic records. While the two institutions running BOOK Plus had some problems in common, system parameters were not uniform and each was on different software releases.

### Integrated Tertiary Software (Cape Technikon)

This was a distributed system operated by Unisys Africa, based on a total software package employed by the institution, with the library module supporting all functions except circulation. Networking was not a component.

### PALS (University of the Western Cape)

This Public Access Library System was implemented in 1986 and upgraded in 1989, with modules added in 1990 when PALS was implemented at the Provincial Library Service. Supported locally by UNIDATA, the system was designed to handle a number of libraries linked to the same server, and so was reputed to be efficient in a networked environment. At UWC it handled all internal library operations, though at the time the software did not support a journal article access system.

### ERUDITE (University of Stellenbosch)

ERUDITE, operated by Universal Knowledge Software (UKS), a subsidiary of UCS Group, was used for all internal library functions including ordering, enquiries, circulation, serials control, and financial administration, and was integrated into the campus network. A major advantage was its efficiency, and the fact that it was the same software

employed at the time by SABINET, the South African Bibliographic Network.<sup>13</sup>

#### 4 The Vision

Where CALICO differs from other consortia in South Africa was in the vision which embraced “the concept of a single Western Cape library collection, that is housed at different locations with all resources accessible to anyone who has need of them.”<sup>14</sup> The collections were to remain in their current locations, but with vastly increased access through a dedicated network linked via metropolitan area services. In this model, the institutions would decide to merge their library operations at a stroke, so that acquisitions, serials management, circulation, and bookkeeping functionality would all operate as if CALICO were a single library.<sup>15</sup> Commitment by all to agreed policies such as cooperative acquisitions and lending, and adherence to agreed standards, would have to be in place to ensure maximum retrieval and also so that all users would assume they were searching a single library collection, even though physical collections would remain the property of the home institution. A factor inherent in the vision to promote information literacy and economic development in the region through information provision was the right of all citizens to access, evaluate and effectively use information to improve their quality of life. Based on shared strength among like institutions, the initial impetus has been with the five tertiary institutions, but more than 300 possible regional beneficiaries were identified, from non-governmental organizations (NGOs) to schools, distance learning centers and the local site of the National Library of South Africa. This indicated a desire to share the

---

<sup>13</sup> See <http://www.erudite.co.za/>.

<sup>14</sup> Patricia Senn-Breivik et al.

<sup>15</sup> *Affirming the CALICO Vision: The Acquisition of the SLIS and its Implications for Policy in Member Libraries: A Discussion Paper* (Cape Town: Project Management Team, October 1997).

burden of undoing the educational inequalities of the past. The cooperative ventures were not seen primarily as money-savers, but rather as ways to increase efficiencies in information and to avoid unnecessary duplication of effort and resources. There was no implicit or explicit attempt for any of the historically disadvantaged institutions to benefit at the expense of the better-endowed; rather, it was understood that together there would be access to a greater range of materials and better services than any one library could provide. More recently, the cost-bearing model has been under scrutiny, due to increased demand on particular libraries from partner institutions.

Investigations looked at practices at each institution and the viability of merging the five databases into a single catalog. In addition to analysis by each institution, two independent consultants who had no vested interest in any one institution also looked at the workflow and standards at each. Three of the institutions cataloged directly onto SABINET<sup>16</sup> and downloaded records, while two cataloged in-house and in theory uploaded their original records and holdings.

## 5 Exchange Format

At the time, the exchange format within the South African library community was SAMARC. This was based on UNIMARC and was developed in the 1970s by order of the National Library Advisory Committee. It was also used as the base format for various commercial and in-house library systems, but this only furthered South African isolation, as it inhibited the exchange of bibliographic records. An investigation into the different MARC formats and a comparison of the costs and benefits of each was mandated by the Interim Committee for Bibliographic Organization (ICBO) and funded by the South African Department of Arts, Culture, Science and Technology.<sup>17</sup>

---

<sup>16</sup> See Pierre Malan, “A National Union Catalog for Shared Cataloging and Resource Sharing by Southern African Libraries,” in this volume.

<sup>17</sup> M. M. M. Snyman, *Investigations into a Future Machine Readable Cataloging (MARC) Format for South African Libraries: A Report with Recommendations*, 1997.

At a seminar held at the University of Pretoria in April 1997 to discuss a future MARC format for South Africa, an overwhelming majority voted to replace SAMARC with USMARC as the preferred format. In January 1999, the MARC Office at the National Library of South Africa conducted the first USMARC interactive online training course.

The only format in SAMARC was for bibliographic records, and there was nothing for authorities, holdings etc. SABINET imported records from various sources (incoming records had to be converted to SAMARC, those from Library of Congress from USMARC, and BNB records firstly from UKMARC to USMARC and then to SAMARC). This resulted in information being dropped along the way, since there were not always corresponding tags and subfields in each format. SABINET had to adopt its own authority file by extracting headings from the bibliographic records, so there was an amount of authority conflict which compromised standards and further inhibited record exchange.

While all libraries in South Africa subscribed to an obligation to contribute to a national catalog for purposes of interlending and collection development, SABINET had been unable to keep abreast of quality control, and had opted for holdings coverage rather than quality of records. The fact that libraries that provided the SACat with records did not all catalog centrally resulted in records of different levels and standards of cataloging, and in a national database where retrieval of records was time-consuming both for cataloging and inter-library loan purposes. SABINET Users Committee had set up an ad hoc committee in 1994 to look at the quality of the database as a whole, but also to investigate the state of the authority files in particular. The University of South Africa (UNISA) investigated subject headings, and the State Library analyzed personal names.<sup>18</sup>

In 1995, Sabinet Online became an international distributor for OCLC, and the CALICO libraries were among the first member libraries in South Africa to use OCLC's PRISM service, though not to its full potential, as some of the systems in use had problems with downloading of long records. An earlier release of BOOK Plus had a record-length restriction, but even

---

<sup>18</sup> Seminar on Bibliographic Standards for Promotion of Cooperation, Pretoria: University of Pretoria, 1996.

when this was increased, particularly long records would not copy to SABINET.

PRISM was useful mainly for music CDs, but SAMARC did not accommodate name title added entries from the 700 tag. This resulted in added entries and variant title tags without any corresponding link. SAMARC used a separate 204 tag for ‘gmd’ (General Material Designation) rather than a subfield in the title field as in USMARC, and this did not convert correctly. SAMARC did not have adequate fixed fields for recordings—printed music translated correctly, but not music CDs, all causing additional manipulation of records before they could be transferred by FTP.

Of the five CALICO institutions, UCT is the only one that uses diacritic codes, mostly for Arabic and Hebrew works. USMARC uses ASCII for special characters, SAMARC had escape sequences, and then BOOK Plus used EBCDIC. Once again this multiplied the potential for error. The tables for conversion had not always been synchronized, and at downloading, the letter with its diacritic might be dropped completely, or else result in a character string remaining embedded in both the bibliographic records and authorities. Once again this caused duplicate headings as well as corruptions in display, filing, and subsequent retrieval.

The University of Cape Town libraries had cataloged on SABINET since 1986, first receiving catalog cards, then a weekly tape and eventually a daily FTP file. When BOOK Plus was installed in 1990, there was no matching program, resulting in duplicate and even multiple occurrences of records loaded from various sources. The basic catalog comprised tapes of bibliographic records that had holdings on SABINET. Some retrospective conversion was done from microfilmed cards by ‘amarc’ Data International in Australia. These records were not upgraded, the data capturers were not catalogers, language problems and distance made quality control difficult so that errors were created and compounded, and the money ran out before the project could be completed. Later retrospective conversion of printed music and scores done by SABINET Special Projects was also problematic, due to lack of music expertise and the poor quality of copy cataloging. The university funded these two projects, since the policy favored outsourcing rather than using data capture on site. CALICO policies favored local

contracting, so in-house retrospective conversion was conducted later for Government publications material and updating of joint serials holdings.

It was necessary for all five institutions to conduct in-depth analysis of their existing catalogs, documenting and prioritizing the clean-up necessary before conversion to USMARC and for possible merging to a single database on the chosen system. This analysis was drawn up in terms of what should be done prior to data extraction, either programmatically or manually, what could be fixed at the time of conversion to USMARC, and what should wait until after implementation or merging, again programmatically and manually. To some degree, what could be done during and after merging was dependent on what system was chosen and what that particular vendor could offer.

Four of the databases required work on headings or authority files; the fifth had no authority module at all, but merely indexes. The ITS database at Cape Technikon filed individual subfields in all author and subject tags in alphabetical order; e.g. a personal name tag having subfields a, q, d, stored them in the order a, d, q. In SAMARC, personal names had an additional subfield 'b' for first name or initials. Correcting the order of subfields was one task identified as having to be done programmatically prior to data extraction.

For various historical reasons, such as importation of records from different sources and embedded or implied punctuation, as well as changes to the length of filing keys, the BOOK Plus database at UCT contained duplicate and variant headings, which resulted in user frustration and possible non-retrieval of relevant items. Cleaning up headings was part of regular cataloging processes, but data corruption caused backlogs and loss of linkage between headings and bibliographic records. Out of more than 667,000 headings, it was estimated by mid-1996 that a possible 40,000 were duplicates, and a further 10% needed maintenance. Professional expertise and familiarity with the history of the database was essential for manual correction.

While Peninsula Technikon also used BOOK Plus, the parameters had been set up differently, and authorities were repeated by type. An author or corporate body might appear in three separate files depending on whether it had primary, alternative or secondary responsibility.

Fixed fields were problematic at all five institutions. UCT had opted to strip them completely from incoming records, and had only limited information in item records from which these could be rebuilt during extraction.

## 6 Connectivity

The implementation of a shared library system presupposes adequate network connectivity. UNINET, which was based on the X25 communications protocol, existed as a higher education telecommunications network that provided a backbone for library cooperation at the national level, but bandwidth was purchased from Telkom, the telecommunications parastatal, which had monopoly status. All of the consortial initiatives within South Africa required adequate and affordable bandwidth, and so a task force was formed to meet with the Minister of Communications and senior Telkom executives. In late 1998, Telkom executives met with a consortium of US donors in New York, and committed themselves to finding a long-term solution to networking requirements as part of Telkom’s contribution to South African development. In turn, the US donors, led by the Mellon Foundation, undertook to fund what became known as the US Donors’ Bandwidth Project for Higher Education in South Africa. A not-for-profit company TENET was established to manage the transition to the new solution.

Meanwhile, Telkom designed a WAN for the CALICO project with 2-MB links between the cooperating institutions. The upfront setup costs and rental for this frame relay network, known as the Adamastor Network, were funded by the Open Society Foundation for South Africa. Maintenance of the main UNIX servers on which both the Production and Development versions of the shared databases are run has been outsourced to Comparex Africa, a local ICT company.

## 7 Choosing the System

The first major step towards the realisation of the CALICO vision was the purchase of a Shared Library Information System. A Project Management Team (PMT) comprising IT, library and management staff was established.

“The brief of the PMT was to interpret and try to convert years of discussion into a proposal for a shared library information system that could be implemented if funding were obtained.”<sup>19</sup> In August 1996, the PMT sent out a formal Request for Funding (RfF) to The Andrew W. Mellon Foundation in the United States. The RfF gave a description of CALICO and its vision and included specifications for the envisaged system and a detailed budget. The Board of the Mellon Foundation approved the RfF and agreed to provide the funds for the purchase and implementation of such a system.

In 1996, the Western Cape Tertiary Institutions Trust sent out a Request for Information (RfI), and this was followed by a Request for Proposal (RfP) to a number of suitable vendors.

On receipt of the response to the RfP from vendors, a short list of three possible systems was decided upon, and arrangements were made for them to host demonstrations in the Western Cape. These took place in April and May 1997. Staff, both academic and library, and students were invited to attend and were requested to complete evaluation forms for each of the systems. The three systems selected for this exercise were ALEPH 500 (Ex Libris), INNOPAC (Innovative Interfaces Inc.) and Virtua (VTLS).

Following the demonstrations, the short list was reduced to two: INNOPAC, and ALEPH 500. A team of librarians and IT staff from the five institutions was selected to travel to Europe and the United States to visit various sites where either one of the two systems was in use. This site visit took place in May 1997.

The next step in the process was for the selection team to evaluate the two systems on the basis of the demonstrations and the site visit report, and to decide on a vendor of choice. A recommendation could then be made to the Western Cape Tertiary Institutions Trust, who would enter into negotiations with the recommended vendor.

---

<sup>19</sup> James Leatt, “Finding a Fit between Vision, Technology and Organization: The CALICO Case,” in *Commonwealth Collaboration in Universities: Three Case Studies*, ed. Alison Rees (London: Association of Commonwealth Universities, 2001).

Given that CALICO was buying a system for the future, it was felt that the ALEPH 500 system met all the given criteria. The following set of criteria was considered in relation to the systems:

### Long-term viability of the product

ALEPH 500 was a technologically advanced system at the beginning of its life cycle. It was seen to be technologically appropriate for CALICO's long-term goals. For example, the system is based on Oracle; it offers multi-tiered client-server architecture; it is possible to use SQL (Sequential Query Language). From the point of view of being able to support a union (merged) catalog, it was important for the chosen system to be able to support a distributed platform, that is, one database living on several computers. A merged, single catalog such as CALICO was planning for a single database. ALEPH 500 was capable of supporting such a distributed platform. Also, ALEPH 500 was acceptable to CALICO from a product strategy perspective: a significantly advanced new-generation product is offered every five years or so, with zero upgrade costs to the user. Another very important feature for this time, ALEPH 500 was Year 2000-compliant.

### Ability of the vendor to cater for a consortial environment

The ALEPH 500 system would be able to offer CALICO patrons a seamless virtual library, and Ex Libris was willing to negotiate with CALICO to meet consortial needs. Ex Libris would be able to convert bibliographic data from SAMARC to USMARC and would be able to merge the data into one catalog.

### Vendor strength

Ex Libris was seen to be an innovative and financially sound company that was growing rapidly in both market strength and in system sales. This opinion has been confirmed in recent years as sales of ALEPH 500 have increased, and it is now used at 700 sites in 50 countries. Most important for library staff was the issue of local support. Ex Libris has a local

distributor based in Cape Town, Avion Information Systems and Services (AVIONISS), that would provide the first level of support.

### Affordability and sustainability

ALEPH 500 was affordable.

### Expansion and outreach

It would be possible to add new member libraries with little difficulty. There was, however, one major cause for concern. The CALICO vision implied very specific requirements for the circulation of material, and the library staff was worried that the ALEPH 500 system would not be able to handle these. As a result of these concerns, circulation experts from CALICO were invited by Ex Libris to go to Israel to meet with staff there and discuss ways of meeting these requirements. This visit took place in December 1997.

This delayed the negotiation process, but by May 1998, both the vendor and the Western Cape Tertiary Institutions Trust, then operating as the Adamastor Trust, were satisfied, and the contract was signed. Plans to implement ALEPH 500 in the five CALICO libraries began.

## 8 Implementation: 1998–1999/2000

To facilitate the implementation of ALEPH 500, a central implementation team, the Pit Crew, was established, comprising Library IT staff with both library functionality and technology expertise. This team was later dissolved and a project manager appointed to oversee the implementation. Each institution also established an in-house implementation group to work with the Pit Crew, and later the Project Manager, during the implementation phase.

The migration of the bibliographic data from the five library systems to one CALICO database involved three distinct activities. Firstly, the data would have to be extracted from the separate databases to a data format (a flat-file); secondly, it would have to be loaded onto the ALEPH system in

this format; and thirdly, the data would have to be merged and then converted from SAMARC to USMARC. This process would ensure that only one bibliographic record per title existed in the shared CALICO bibliographic file. The literature on the subject of merging identifies three options for handling records identified as duplicates:

1. “One record is chosen as the master record and the others are deleted.
2. All records are kept but clustered around a master record.
3. One record is chosen as the master record and variant fields from the duplicates are added to the master.”<sup>20</sup>

Option one would involve deciding on selection criteria, such as always keeping the record with the highest encoding level, or always keeping the record ‘belonging’ to the institution perceived to have the highest quality records, but would necessarily result in some libraries losing data that they considered valuable and useful to their patrons. Consequently, the preferred option for CALICO was the third one.

## 9 The ‘Big Bang’ Approach

The plan for implementation at this stage was that all institutions would go live at very much the same time: the big bang approach. In other words, the bibliographic data would be extracted from all five institutions. The first database would then be loaded onto the ALEPH 500 platform; then the second would be loaded and the data matched and merged with the first; then the third would be loaded, matched and merged with the now combined first and second databases, etc., until all five databases had been loaded, matched and merged. Ex Libris would provide the matching algorithm, and the conversion from SAMARC to USMARC would be carried out by AVIONISS, using specifications drawn up by and purchased from Sabinet Online.

---

<sup>20</sup> Susan S. Lazing, “To Merge or Not to Merge: Israel’s Union List of Monographs in the Context of Merging Algorithms,” *Information Technology and Libraries*, 13/3 (1994).

While the systems librarians and implementation teams were working on the implementation, the Regional Catalog of Monographs Committee (RCMC) was investigating the feasibility of the CALICO vision of a merged database. The RCMC was established in 1995, prior to the selection of the system and the signing of the contract, and was composed of catalogers from the five libraries. The mandate of this committee was to investigate the possibility of creating a union catalog for the Western Cape. To be able to do this, they would need to establish whether it would in fact be possible to merge the databases, given that the institutions were using different systems and that there were disparities in the cataloging practices followed by each institution. The other option was for each library to maintain its own database separately on a shared automated system, but this would conflict with the overall vision of CALICO.

They were also mandated to investigate methods of overcoming the use of different systems and local practices and to develop and implement a cooperative cataloging program between the five libraries.

To facilitate the making of these decisions, two independent consultants were asked to investigate the differences in the catalogs of the five libraries, establish which cataloging standards were used and identify the differences in the interpretation of these standards, and to make a recommendation on the basis of their belief concerning the possibilities of merging the databases of the five libraries into one.<sup>21</sup> The reports of the consultants were submitted to the RCMC and highlighted a number of issues that could impede the formation of a merged catalog.

---

<sup>21</sup> Karin De Jager, "Cape Library Cooperative, Regional Catalogue of Monographs Investigation: Report," unpublished paper (Cape Town: University of Cape Town, 1996); Ziska Johnson, "Investigation into the Differences between the Five CALICO Libraries and the Feasibility of Forming an Amalgamated Catalogue," unpublished paper (Stellenbosch University, 1996).

## 10 Deviations from Standards, and Local Practices

Although all the libraries adhered to international standards like AACR2 and Library of Congress Subject Headings (LCSH), there were deviations in some instances.

### University of Cape Town

Subject headings from an in-house African Studies thesaurus had been added to records using the standard SAMARC 600–607 subject headings tags. The 650 SAMARC tag for keywords was not used, and no subfield code had been used to identify these headings as being from a source other than LCSH. SAMARC allowed a subfield \$2 to identify the source of the headings.

### MeSH (Medical Subject Headings)

Subject headings were used on titles belonging to the Medical Library, and like the African Studies headings, the source subfield \$2 had not been used, so there was no easy way to isolate these headings.

Staff in the Library’s Reserve section added very basic records to the database, for material such as photocopies and articles that had been placed on short loan. This was necessary so that the material could be circulated online. Typically, these records would not have any subject analysis or publication and imprint details. Thus, there were records of varying levels of completeness.

### Peninsula Technikon

Library of Congress Subject Headings had only been used since 1993. From 1989 to 1993, SEARS had been used, and prior to that, free language subject headings had been applied.

Titles in Afrikaans had the Afrikaans version of the corporate heading, and English titles used the English version.

The way in which headings for Government departments had been structured was inconsistent: there were entries under the name of the

country and then the department, and also entries directly under the name of the department.

### Cape Technikon

As at the Peninsula Technikon library, both English and Afrikaans subject headings were used, depending on the language of the publication. Free text indexing terms were used in the SAMARC 650 keyword tag.

### University of the Western Cape

Cataloging standards had not been applied uniformly, since the Library did not have a Cataloging Department. Cataloging was the responsibility of the subject librarians, along with their reference and other duties, and there was no quality control or checking of their cataloging work.

Although Library of Congress Subject Headings were used, they were not uniformly applied, and the *Thesaurus of South African Socio-Political and Economic Terms from an Anti-Apartheid Perspective*<sup>22</sup> was also used for titles with a specific South African content, which were not covered in sufficient detail in LCSH. Like the University of Cape Town, these non-LCSH headings were added in the standard SAMARC 600–607 tag and not in a keyword tag. British spelling was used in the subject headings, and not the standard American. ‘South Africa’ was routinely dropped from the headings for South African government departments and bodies, and all were entered under ‘Department’. Headings for personal names were not consistent. If the title of the work was in both English and Afrikaans, then both versions of the name were given, if the title was in Afrikaans, then the Afrikaans version was used and the English version for a work published in English.

---

<sup>22</sup> Christopher Merrett, *Thesaurus of South African Socio-political and Economic Terms from an Anti-Apartheid Perspective*, 2nd ed. (Pietermaritzburg: University of Natal Library, 1991).

## University of Stellenbosch

This Library had a very high standard of cataloging, and there were few deviations from the recognized standards. Only authorized subject headings from the latest version of Library of Congress Subject Headings were used, and only authorized forms of personal names were used. English forms of corporate names were used if the work was in English, and Afrikaans if the work was in Afrikaans.

### 11 Authority Files

All of the libraries, with the exception of the University of the Western Cape, had authority or headings files. All reported duplicate headings, except for the University of Stellenbosch. In fact, the University of Stellenbosch was the only institution with an authority file that was well maintained, with few duplicates and deviations.

### 12 Staff Perceptions

The consultants interviewed staff in the Cataloging Departments of the Libraries and asked whether they thought a single catalog was feasible. It became apparent that not all staff thought such a union catalog was possible, and many were in “favour of retaining control over their own databases and simply providing bibliographic access to the other members of CALICO.”<sup>23</sup>

The primary concern was to maintain the integrity of the database. Staff felt that it would be difficult to maintain high standards of cataloging, primarily because of the differing levels of expertise of those who would have cataloging rights in the database. For example, cataloging of material at the University of the Western Cape was the responsibility of the reference librarians, and not specialist catalogers. Similarly, Circulation and Acquisitions staff added brief records to the University of Cape Town

---

<sup>23</sup> De Jager.

database. It was felt that the local cataloging practices of each institution would affect the success of the merge and would compromise the quality of the database.

### 13 Recommendations

The consultants came to the conclusion that it would be possible to merge the five databases, but a number of conditions would have to be met. Primarily, mutually acceptable standards would have to be negotiated and established. Individual cataloging practices would have to be standardized and minimum levels agreed to. Most importantly, all the institutions would have to commit themselves to adhering to these standards. It was also recommended that certain staff should take responsibility for quality control of the database. It was agreed that maintaining the merged catalog would be expensive and time-consuming, and CALICO would have to ensure that sufficient resources, both financial and staff, were made available.

Following the release of these reports, the RCMC began to work on two projects that would help to standardize cataloging practices and ensure that an acceptable standard of cataloging was maintained in the merged catalog. The first of these was the compilation of a Cataloging Procedures Manual which would establish cataloging standards and processes, and the second was the formulation of core records for all formats. These were completed in 1999. In addition to working on the above projects, each institution undertook major clean-up projects on their own databases prior to merging.

While the RCMC was preparing the catalogs for the final extraction and merge, a major problem had arisen at the University of the Western Cape. The PALS system was being run on hardware that was becoming increasingly expensive to maintain, and it no longer made sense for the institution to continue to pay for its maintenance when a new system had been purchased and implementation was imminent. The Project Management Team and Ex Libris discussed the possibility of a staggered implementation, with UWC converting to ALEPH 500 ahead of the other four institutions.

As a result of UWC's problems, it was agreed that the 'big bang' approach to implementation would not be followed, but that each institution would go live at different times during the course of 1999. This had major implications for the proposed merging of the catalogs. Since implementation was to be staggered, all stages of implementation would, therefore, also be staggered. Extraction and loading of the data would take place individually prior to each institution going live, and since CALICO was committed to working on ALEPH 500 in USMARC format, conversion from SAMARC to USMARC would also have to take place prior to each institution going live. The merging of the five databases could now take place only after all the institutions had gone live on ALEPH 500, and would be the final step in the implementation process.

## 14 Implementation Schedule

The implementation of ALEPH 500 in the CALICO libraries took place according to the following schedule:

- February 1999: the University of the Western Cape went live with all modules on version 11.5;
- March 1999: the Cape Technikon went live with the cataloging and acquisitions modules, also using version 11.5;
- May 1999: the University of Cape Town went live with the cataloging module, again on version 11.5;
- Version 12.1 became available in July/August 1999, and the three 'live' institutions upgraded. It was hoped that functionality needed by the CALICO libraries that was missing in version 11.5 would now be available;
- July 1999: the Cape Technikon implemented the circulation module and the Web OPAC.

UCT had planned to implement the circulation module with version 12.1, but staff were still not satisfied that they would be able to offer their patrons the same level of service as they had using BOOK Plus. Consequently, it was decided to wait for the next version, 12.2, that would allow them to offer an equivalent service.

- October 1999: the Peninsula Technikon implemented all modules with the exception of acquisitions;
- November 1999: version 12.2 was ready for installation;
- November 1999: University of Cape Town implemented the circulation module and the Web OPAC, and Peninsula Technikon implemented the acquisitions module using the new version.

It was imperative that ALEPH 500 be implemented at all the institutions prior to January 2000, as the old systems were not Y2K compliant. Since Stellenbosch University ran a fully integrated system on ERUDITE, it was important for it to wait until year-end before extracting and converting the financial transactions from ERUDITE to ALEPH 500.

- December 1999: the University of Stellenbosch implemented the cataloging and acquisition modules;
- January 2000: it implemented the circulation module, Web OPAC, the serials module and all financial transactions.

## 15 Post-Implementation 2000–2001

Thus, by January 2000, the five CALICO institutions were all using the same library information system, ALEPH 500 version 12.2, and all were running the system from the same server. CALICO was still a long way from realizing the original vision of a shared union catalog, and in reality it appeared as if CALICO had become “a mere aggregate of current library practice,”<sup>24</sup> precisely what the Project Management Team did not want to happen.

Following implementation, each institution configured the system to meet its own institution’s specific needs. For example, each institution still followed its own circulation practices, and the system was configured to meet these specific needs and rules, despite the fact that some attempts had been made by the institutions prior to implementation, to standardize these and to agree on common parameters within the table set up. There is no

---

<sup>24</sup> *Affirming the CALICO Vision...*

shared circulation system, and patrons do not have equal access rights on all campuses. Each institution has assumed responsibility for maintaining its own database, and although there are common practices and standards which are being adhered to more stringently, it is not the shared catalog of the CALICO vision.

The users of the institutions are able to search the collections of the other institutions, but they do not have equal access rights on all campuses. They are able to borrow material from the other libraries, but this is organized through the inter-library loan departments of the institutions.

Does this mean that the original “bold vision of a shared library information system—a ‘library without walls’” or a “‘single, pooled library system’ that would link collections housed separately,”<sup>25</sup> has been abandoned? No, it does not, but where does CALICO go from here to meet the vision?

The face of the modern library has undergone rapid changes in the last few years and each of the five CALICO libraries has been affected by these changes. The traditional print collections have been supplemented and enhanced by the online database industry that has “enabled libraries to provide access to additional resources that are not necessarily owned by the library.”<sup>26</sup> Changes in the format and in the source of information from traditional print to electronic has led inevitably to changes in the needs of the library users and the CALICO libraries found themselves facing a new challenge. To summarize: libraries and the vendors of library systems need to find a technological solution that will “support robust integration of locally-held information resources with licensed networked databases and with Internet-based resources.”<sup>27</sup>

A number of library vendors have met this challenge and have developed new services and software that will allow end-users access to both print collections and electronic resources. As users of the ALEPH 500

---

<sup>25</sup> Leatt.

<sup>26</sup> Tom Kochtanek, “New Developments in Integrated Library Systems,”  
<http://gessler.emeraldinsight.com/vl=2844020/cl=23/nw=1/rpsv/librarylink/technology/nov01.htm>.

<sup>27</sup> Kochtanek.

system, it was inevitable that CALICO would investigate the two new products developed by Ex Libris: MetaLib and SFX. “MetaLib is the perfect platform for managing a hybrid library environment, including both the emerging electronic collection with its digital resources and the traditional library with its print resources. MetaLib serves as a gateway to local and remote databases.”<sup>28</sup> “SFX is a unique and revolutionary tool for navigation and discovery, delivering powerful linking services in the scholarly information environment. With SFX libraries can define rules that allow SFX to dynamically create links that fully integrate their information resources regardless of who hosts them – the library itself, or external information providers.”<sup>29</sup> Given the new technology available, at its meeting on March 28<sup>th</sup>, 2001, the CALICO Management Committee “agreed that CALICO would purchase from remaining Mellon funding one version of MetaLib/SFX.”<sup>30</sup> At the following meeting held a month later on April 26<sup>th</sup>, “[i]t was formally noted that the decision to purchase MetaLib/SFX software replaced the original decision to merge the catalogs of the five institutions.”<sup>31</sup> Upgrading to ALEPH 500 Version 14.2 took place in mid-2002, and implementation of MetaLib and SFX at the five institutions will begin early in 2003.<sup>32</sup>

## 16 Conclusion

It has been a long journey from the initial vision of 1992 to the current situation, and ‘over-democratization’ can be seen as having been a major

---

<sup>28</sup> See <http://www.exlibris.co.il/metalib/index.html>.

<sup>29</sup> See <http://www.sfxit.com/body.html>.

<sup>30</sup> Minutes of the 13th meeting CALICO Management Committee, Wednesday 28th March 2001, Boardroom, University of Stellenbosch Business School, point 11.

<sup>31</sup> Minutes of the 14th meeting CALICO Management Committee, April 26, 2001, UWC Council Chambers, Bellville.

<sup>32</sup> For implementations of MetaLib and SFX, see also Bohdana Stoklasová and Pavel Krbec, “CASLIN Uniform Information Gateway,” in this volume.

deterrent to the implementation of the CALICO vision. The effort to involve as many people as possible in the decision-making process resulted in a multitude of committees and working groups and in time-consuming institutional consultation, where "CALICO became hostage to the veto of one, and often found itself going at the pace of the slowest."<sup>33</sup>

Developments in technology have overtaken the original plan to merge the five catalogs. With the purchase of MetaLib and SFX, CALICO will achieve a virtual union catalog without having to go through the complexities of physically merging the five databases. The difficulties of actually merging the databases, with their different standards of cataloging and idiosyncratic practices, and the time-consuming checking of a merging algorithm have been avoided. Advances in technology will in fact offer far more than the original vision: a CALICO user will be able to search the library catalogs of all five institutions, as well as being able to access the libraries' electronic resources at the same time via a single gateway. Patrons will have to familiarize themselves with only one interface, but will have a wealth of scientific and scholarly information available to them, both in print and electronic format.

## **Appendix: Acronyms and Abbreviations**

ACE	American Council of Education
AVIONISS	Avion Information Systems and Services
BTech	Bachelor of Technology
CALICO	Cape Library Cooperative
CTK	Cape Technikon
ICBO	Interim Committee for Bibliographic Organization
ITS	Integrated Tertiary Software

---

<sup>33</sup> Leatt.

NGO	Non-Governmental Organization
OCLC	Online Computer Library Center
PALS	Public Access Library System
PTK	Peninsula Technikon
RCMC	Regional Catalog of Monographs Committee
SABINET	South African Bibliographic Network
SACat	South African Catalog
SAMARC	South African MARC format
SQL	Structured Query Language
TENET	Tertiary Education Network
UCT	University of Cape Town
UWC	University of the Western Cape

## Contributors

*Janne Andresoo* has an M.A. in Information Science and is responsible at the National Library of Estonia for the coordination of Estonian bibliographic activities and the implementation of standards related to cataloguing. The Library leads the Working Group on Cataloguing of the Estonian Libraries Network Consortium (ELNET), coordinates the generation of the electronic catalogue ESTER for thirteen Estonian libraries and contributes to the teaching of subjects related to information management at Tallinn Pedagogical University.

*Alojz Androvič*, Ph.D. Ing., is the Deputy Director of the University Library, Bratislava, Slovakia and the director of the ISSN National Centre of Slovakia. He has also been responsible for the CASLIN Digital Library Portal project at the University Library and is the author of several articles on library automation.

*Géza Bakonyi* is the Deputy Director General of the University Library, Szeged, Hungary. He is also the Head of the Library and Information Science Department and has been Project Manager of the Hungarian Shared Cataloging Program since 2001.

*Nadia Caidi* received her Ph.D. from the University of California, Los Angeles (UCLA), and is now teaching in the Faculty of Information Studies at the University of Toronto. Her research interests are information policy, social and community informatics, and the role of culture in information seeking, design, and use.

*Rein van Charldorp* is the Managing Director of OCLC PICA BV in Leiden, The Netherlands. OCLC PICA is owned by OCLC and Stichting PICA and is responsible for Europe, Middle East and Africa. Before joining OCLC PICA he worked at Reed Elsevier in both the science and business divisions. In the last 10 years with Reed Elsevier, he held management positions in Amsterdam, New York, Boston and Ireland, co-directing the migration of the print products into Science Direct.

*Karen Coyle* is a librarian with nearly thirty years' experience in digital libraries. She was a developer on the University of California's MELVYL

union catalog from the early 1980s. Her career interests range from technical standards to social policy. Her website is <http://www.kcoyle.net>.

*Lettie Erasmus* is a Deputy Director at the University of South Africa Library with specific responsibility as Library System Coordinator. From 1996 to 2000 she served as a member of the GAELIC System Implementation Team, and from 2001 has been the Leader of the GAELIC INNOPAC / Millennium System Workgroup.

*Błażej Feret* is a Deputy Director of the Main Library of the Technical University of Łódź, and System Librarian for the Łódź Academic Library Network, a consortium of 10 academic libraries in the city of Łódź, Poland. Since 1996, he has been responsible for The Andrew W. Mellon Foundation-supported purchase and implementation of the Horizon library system in Łódź library consortium.

*Janifer Gatenby* is a consultant at OCLC PICA. Since 1983 she has worked actively with interoperability standards and is the editor of ISO 8459, parts 4 and 5 (data elements), the Union Catalog profile and a contributor to the ZING SRU standard, the Bath Profile and Z39.50 Holdings schema. She is currently involved with a working group on FRBR and Continuing Resources.

*Stefan Gradmann* has, since 2000, been the head of Hamburg University's Information Management unit, which is part of the computing center and whose mission includes e-publication services and providing access to electronic information resources. Earlier, he worked for Pica in the Netherlands as product manager (1997–2000) and directed a large library cataloging network (GBV, 1993–1997).

*Henryk Hollender* is responsible for projects and cooperation at the Warsaw University Library. Previously a library science instructor, he was director of the Warsaw University Library from 1992 to 2003.

*Ole Husby* is a senior advisor at BIBSYS, the Norwegian library network for academic and special libraries. He is presently head of the development department, and is participating in the national effort of building a Norwegian Digital Library.

*Annu Jauhiainen* is the Deputy Director of IT at Helsinki University Library, the National Library of Finland. She is the Project Manager and Coordinator of the Linnea2 Consortium, which consists of all Finnish University Libraries, the National Library and some special libraries.

*Klára Koltay* is an assistant director of the University and National Library at Debrecen University. She has been in charge of building the library's database and of subject cataloging programs since 1994, and has been involved in Hungarian cooperative cataloging projects.

*Pavel Krbec* has been head of the Charles University Computer Science Center since 1991. He is responsible for Prague Academic Computer Network and for Charles University's information system, including library automation.

*Gabriela Krčmařová* graduated from the Institute of Library and Information Science of Charles University in Prague. From 1981 to 2002, she worked in the National Library of the Czech Republic. In 1994, she was appointed head of the CASLIN Union Catalog Department. She was responsible for contracts with the Union Catalog participants and for the procurement of technology and software. Since May 2002, she has been working at CESNET (Czech National Research and Education Network).

*Andrew Lass* is Professor of Social and Cultural Anthropology at Mount Holyoke College, MA., and the project manager of the Mellon funded CASLIN projects and seminars. He is the author of several articles on library modernization in the Czech and Slovak republics and co-editor, with Richard E. Quandt, of *Library Automation in Transitional Societies* (Oxford, 2000).

*Martin Lhoták* is the head of IT department in The Library of The Academy of Sciences, Czech Republic. He has been responsible for the union catalog of the Academy since 1998. At present, he is working on the implementation of a new library information system in the Academy.

*Pierre Malan* is the Director of the Academic and Library Division of Sabinet Online Ltd. He has been extensively involved in the implementation of automation systems in South African library consortia since 1996, and has

been overseeing the upgrading of the South African National Union Catalog since 1998.

*Dianne Leong Man* is Deputy University Librarian (Technical Services) at the University of the Witwatersrand. Since the inception of GAELIC in 1996, she has served as a member of the GAELIC System Implementation Team, then as Leader of the GAELIC INNOPAC Workgroup and currently as Leader of the GAELIC Information and Communications Technology Team.

*Amanda Noble* is a Senior Librarian in Bibliographic Access at the University of Cape Town Libraries. From 1998 to 2000, she was seconded to the team responsible for the implementation of the ALEPH 500 system at the University of Cape Town. Currently she is responsible for cataloging electronic resources and serials for the Library and maintaining the ALEPH cataloging and authority file tables, and serves on the Sabinet Online Standards Committee.

*Riin Olonen* is an Executive Director of the Estonian Libraries Network (ELNET) Consortium. She is also the Library System Manager at the National Library of Estonia.

*Andrzej Padziński* is a Deputy Director of the Main Library of the University of Agriculture in Lublin, Poland.

*Anna Paluszkiewicz*, now deceased, worked at the NUKat Center (National Union Catalog Center) of the Warsaw University Library. She was responsible for cataloging standards and database consistency.

*Richard E. Quandt* is Emeritus Professor of Economics at Princeton University and a senior advisor at The Andrew W. Mellon Foundation. Since 1990, he has directed the Foundation's East European program, and from 1994 to 1999 he co-directed the Foundation's program on digital libraries.

*Norma Read* is manager of Bibliographic Access at the University of Cape Town Libraries, and a member of the Libraries Management Team. Since its inception in 1999, she has been a member of the Sabinet Online Standards Committee, and is currently (2003–4) the transitional representative for South Africa on OCLC Members Council.

*Lýdia Sedláčková*, Ph.D., is Head of the Union Catalogue Department of the University Library, Bratislava, Slovakia. She is the author of several articles on union cataloguing in Slovakia and co-author of the project funded by the Open Society Fund to develop a Slovak union catalogue of periodicals.

*Bohdana Stoklasová* is a Deputy for Library Management at the National Library of the Czech Republic. She is responsible for several national projects in retrospective conversion of catalogs and library automation. She is the Chair of the IFLA Bibliography Section.

*Ilona Trtiková* is a student at the Institute of Library and Information Science of Charles University in Prague. She worked in the National Library of the Czech Republic as a team member at the Czech national union catalog CASLIN (administration and development) until October 2002. Since November 2002, she has been working at the Czech Technical University. Her professional interests include union catalogs, library software and Web information architecture.

*Erik I. Vajda* worked extensively in various fields of information retrieval, standardization, abstracting services, etc., and eventually became the founder and first manager of the shared cataloguing system of Hungarian technical libraries. He was also the project manager of the MOKKA project establishing a Hungarian national shared cataloguing system.

*Tomasz Wolniewicz* is the Director of Information and Communication Services Center at the Nicolaus Copernicus University, Toruń, Poland. From 1999 to 2002 he was the deputy director of the University Library, responsible for computerization. He has developed the Distributed Catalog of Polish Libraries, KaRo.



## Conference Participants

<b>Name</b>	<b>Address</b>	<b>Tel./fax/e-mail</b>
Aldis Abele	Ventspils muzejs 3 Akmenu Ventspils 3601 Latvia	Tel.: 371-362-2031 Fax: 371-920-7303 abele@vm.apollo.lv
Janne Andresoo	Eesti Rahvusraamatukogu Tõnismägi 2 15189 Tallinn Estonia	Tel.: 372-630-7104 Fax: 372-631-1410 janne@nlib.ee
Alojz Androvič	Univerzitná knižnica Michalská 1 814 17 Bratislava Slovakia	Tel.: 421-2-5443-3246 Fax: 421-2-5443-3754 androvic@ulib.sk
Géza Bakonyi	University Library Szeged University Dugonics tér 13 6720 Szeged Hungary	Tel.: 36-62-544-039 Fax: 36-62-544-035 bakonyi@bibl.u-szeged.hu
Voitěch Balík	Národní knihovna Klementinum 190 110 01 Prague Czech Republic	Tel.: 420-2-2166-3262 Fax: 420-2-2166-3261 vojtech.balik@nkp.cz
Jack Bazuzi	VTLS, Inc. 1701 Kraft Drive Blacksburg, VA 24060 USA	Tel.: 540-557-1200 Fax: 540-557-1210 bazuzij@vtls.com
Pēteris Briedis	Latvian Academic Library Rūpniecības ielā 10 Riga 1235 Latvia	Tel.: 371-710-6217 Fax: 371-710-6202 peteris.briedis@lib.acadlib.lv

<b>Name</b>	<b>Address</b>	<b>Tel./fax/e-mail</b>
Eva Bulinová	Computer Science Center Library Department Charles University Ovocný trh 3-55 116 36 Prague Czech Republic	Tel.: 420-2-2449-1238 Fax: 420-2-2449-1588 Eva.Bulinova@ruk.cuni.cz
Maria Burchard	Biblioteka Uniwersytecka w Warszawie l. Dobra 56/66 00 312 Warsaw, Poland	Tel.: 48-22-552-5350 Fax: 48-22- 552-5659 buwmb@mercury.ci.uw.edu.pl
Nadia Caidi	Faculty of Information Studies University of Toronto 140 St. George Street Toronto, ONT, M5S 3G6 Canada	Tel.: 416-978-4664 Fax: 416-971-1399 caidi@fis.utoronto.ca
Rein van Charldorp	OCLC PICA Schipholweg 99 PO Box 876 2300 AW Leiden The Netherlands	Tel.: 31-71-524-6500 Fax: 31-71-522-3119 r.vancharldorp@oclcpic.nl
Ewa Chrzan	University Library University of Gdańsk ul. Armii Krajowej 110 81-824 Sopot Poland	Tel.: 48-58-551-5221 Fax: 48-58-551-5221 ewa@panda.bg.univ.gda.pl
Karen Coyle	California Digital Library 1111 Franklin Street Oakland, CA 94607-5200 USA	Tel.: 510-987-0567 kcoyle@kcoyle.net
Ewa Dobrzyńska- Lankosz	University of Mining and Metallurgy Al. Mickiewicza 30 30-059 Kraków Poland	Tel.: 48-12-637-3208 Fax: 48-12-634-1404 lankosz@bg.agh.edu.pl

<b>Name</b>	<b>Address</b>	<b>Tel./fax/e-mail</b>
Helena Dvořáková	Národní knihovna Klementinum 190 110 01 Prague 1 Czech Republic	Tel.: 420-2-2166-3385 Fax: 420-2-2166-3174 helena.dvorakova@nkp.cz
Maria Dzaviková	Univerzitná knižnica v Bratislave Michalská 1 814 17 Bratislava Slovakia	Tel.: 421-7-5443-1151 Fax: 421-7-5443-4246 maja@ulib.sk
Lettie Erasmus	UNISA Library PO Box 293 Pretoria 0003 South Africa	Tel.: 27-12-429-3342 Fax: 27-12-429-2925 erasmaj@unisa.ac.za
Błażej Feret	Main Library Technical University ul. Zwirki 36 90-924 Łódź Poland	Tel.: 48-42-312056 Fax: 48-42-363165 blzferet@sunlib.p.lodz.pl
Janifer Gatenby	OCLC PICA Schipholweg 99 2300 AW Leiden The Netherlands	Tel.: 31 71 524 65 00 Fax: 31 71 522 31 19 janifer.gatenby@oclc-pica.org
Anita Goldberga	National Library of Latvia K. Barona iela 14 1423 Riga Latvia	Tel.: 371-722-5135 Fax: 371-728-0851 anitag@lbi.lnb.lv
Mirosław Górny	Poznańska Fundacja Biblioteka Naukowych ul. Powstanców Wielkopolskich 16 61-895 Poznań Poland	Tel.: 48-61-854-3141 Fax: 48-61-854-3149 mgorny@hum.amu.edu.pl

<b>Name</b>	<b>Address</b>	<b>Tel./fax/e-mail</b>
Elżbieta Górską	Biblioteka Publiczna miasta stołecznoġo Warszawy ul. Koszykova 26/28 00-950 Warsaw Poland	Tel.: 48-22-621-7852 Fax: 48-22-621-1968 elzbieta.gorska@biblpubl.waw.pl elzbieta.gorska@sun1.biblpubl.waw.pl
Stefan Gradmann	Projekt Virtuelle Bibliothek Regionales Rechenzentrum der Universitaet Hamburg Schlueterstrasse 70 20146 Hamburg Germany	Tel.: 49-40-42838-3093 Fax: 49-40-42838-6270 stefan.gradmann@rrz.uni-hamburg.de
Esko Häkli	Universitetsbibliotek PB 15 Unionsgatan 36 SF-00014 Helsingfors Universitet Finland	Tel.: 358-9-1912-2721 Fax: 358-9-1912-2719 esko.hakli@helsinki.fi
Henryk Hollender	Biblioteka Uniwersytecka w Warszawie ul. Dobra 56/66 00 312 Warsaw Poland	Tel.: 48-22-552-5663 Fax: 48-22- 552-5659 h.hollender@uw.edu.pl
Aleš Horák	Univerzita Palackého Informační centrum Biskupské nám. 1 771 47 Olomouc Czech Republic	Tel.: 420-68-593-1756 Fax: 420-68-523-5276 ales@rupnw.upol.cz
Ádám Horváth	Országos Széchényi Könyvtár Budapest Várpalota "F" Épület 1827 Budapest Hungary	Tel.: 36-1-224-3737 Fax: 36-1-375-6167 adam@oszk.hu
Ole Husby	BIBSYS 7491 Trondheim Norway	Tel.: 47-7359-2959/47-9117-4086 ole.husby@bibsys.no

<b>Name</b>	<b>Address</b>	<b>Tel./fax/e-mail</b>
Ivars Indāns	National Library K. Barona ielā 14 Riga, 1423 Latvia	Tel.: 371-728-1640 Fax: 371-962-9379 Cell: 371-962-9379 iindans@lbi.lnb.lv
Aija Janbicka	Scientific Library of the Riga Technical University Kipsalas 10 Riga 1658 Latvia	Tel.: 371-708-9443 Fax: 371-708-9474 rtusclib@acad.latnet.lv
Jüri Järs	Technical University Library Akadeemia tee 7 0027 Tallinn Estonia	Tel.: 372-620-3564 Fax: 372-620-3561 jars@lib.ttu.ee
Annu Jauhiainen	Helsinki University Library PO Box 26 (Teollisuuskatu 23) 00014 University of Helsinki Finland	Tel.: 358-9-191-44296 Fax: 358-9-753-9514 annu.jauhiainen@helsinki.fi
Ivana Kadlecová	Library of the Czech Academy of Science Národní 3 115 22 Prague 1 Czech Republic	Tel.: 420-2-2424-0524 Fax: 420-2-2424-0611 kadlec@lib.cas.cz
Mari Kannusaar	Eesti Rahvusraamatukogu Tõnismägi 2 15189 Tallinn Estonia	Tel.: 372-630-7503 Fax: 372-631-1410 mari@nlib.ee
Dušan Katusčák	SNK Nám. J. C. Hronského 1 036 01 Martin Slovakia	Tel.: 421-43-423-7407 Fax: 421-43-430-1802 katuscak@snk.sk

<b>Name</b>	<b>Address</b>	<b>Tel./fax/e-mail</b>
Venta Kocere	Latvia Academic Library Rūpniecības ielā 10 1235 Riga Latvia	Tel.: 371-710-6206 Fax: 371-710-6202 vkocere@lib.acadlib.lv
Klára Koltay	Central University Library Debrecen University PO Box 39 4010 Debrecen Hungary	Tel.: 36-52-410443 Fax: 36-52-410443 kkoltay@giant.lib.klte.hu
Darina Kožuchová	Univerzitná knižnica UPJŠ, Garbiarska 14 042 07 Košice Slovakia	Tel.: 421-55-633-1080 Fax: 421-55-633-1080 kozuchov@kosice.upjs.sk
Pavel Krbec	Computer Science Center Charles University Ovocný trh 3 110 00 Prague 1 Czech Republic	Tel.: 420-2-2449-1232 Fax: 420-2-2449-1588 krbec@cuni.cz
Gabriela Krčmařová	Útvar podpory členů a zákazníků CESNET z.s.p.o. Žitkova 4 160 00 Praha 6 Czech Republic	Tel.: 420-2-2435-2916 Fax: 420-2-2435-5286 gabriela.krmarova@cesnet.cz
Andrew Lass	Department of Anthropology Mount Holyoke College South Hadley, MA 01075 USA	Tel.: 413-533-4721 Fax: 413-536-0263 alass@mtholyoke.edu
Toomas Liivamägi	Tartu University Library 1 W. Struve Tartu Estonia	Tel.: 372-737-5700 Fax: 372-737-5701 Cell: 372-51-53-274 toomas@utlib.ee

<b>Name</b>	<b>Address</b>	<b>Tel./fax/e-mail</b>
Martin Lhoták	Library of the Czech Academy of Sciences Národní třída 3 115 22 Prague Czech Republic	Tel.: 420-2-2140-3471 Fax: 420-2-2424-0526 lhotak@lib.cas.cz
Danuše Lošťáková	Informační centrum Univerzita Palackého Biskupské nám 1 771 41 Olomouc Czech Republic	Tel.: 420-68-563-1756 Fax: 420-68-593-1756 lostakova@rucs.upol.cz
Béla Mader	University Library Szeged University Dugonics tér 13 6720 Szeged Hungary	Tel.: 36-62-544-036 Fax: 36-62-544-035 mader@bibl.u-szeged.hu
Jarmila Majerova	Slovak National Library Nam. J. C. Hronského 036 01 Martin Slovakia	Tel.: 421-43-430-1803 Fax: 421- 43-430-1802 majerova@snk.sk
Pierre Malan	Sabinet Online Bag 9785 Centurion 0046 South Africa	Tel.: 27-12-663-4954 Fax: 27-12-663-3543 pierre@info1.sabinet.co.za
D. L. Man	University of Witwatersrand Private Bag X1 WITS 2050 South Africa	Tel.: 27-11-716-3816 Fax: 27-11-403-1421 man.d@library.wits.ac.za
Adam Manikowski	al. Jana Pawła II 61 m.68 01-031 Warsaw Poland	Tel.: 48-22-636-6326 hmanik@obta.uw.edu.pl hmanik@supermedia.pl

<b>Name</b>	<b>Address</b>	<b>Tel./fax/e-mail</b>
Margarita Marcinkeviča	Library Information Network Consortium 75 Tērbatas ielā 1001 Riga Latvia	Tel.: 371-731-2790 Fax: 371-731-2790 martina@latnet.lv
Küllli Moont	University of Tartu W. Struve 1 50091 Tartu Estonia	Tel.: 372-737-5711 Fax: 372-737-5701 kyllim@utlib.ee
Jan Andrzej Nikisch	Poznańska Fundacja Bibliotek Naukowych ul. Powstancow Wielkopolskich 16 61-895 Poznań Poland	Tel.: 48-61-854-3141 Fax: 48-61-854-3149 nikisch@pfsi.poznan.pl
Amanda Noble	University of Cape Town Private Bag Rondebosch 7701 South Africa	Tel.: 27-21-650-3111 Fax: 27-21-689-7568 mandy@utlib.uct.ac.za
Thomas I. Nygren	The Andrew W. Mellon Foundation 282 Alexander St Princeton, NJ 08540 USA	Tel.: 609-924-9424 Fax: 609-683-4471 tin@mellon.org
Riin Olonen	Eesti Rahvusraamatukogu Tõnismägi 2 15189 Tallinn Estonia	Tel.: 372-630-7189 Fax: 372-630-1410 riin@nlib.ee
Reet Oruste	Estonian Academic Library 10 Rävåla Avenue 15042 Tallinn Estonia	Tel.: 372-665-9409 Fax: 372-665-9400 reet@ear.ee

Name	Address	Tel./fax/e-mail
Sandra Ozolina	National Library of Latvia K.Barona ielā 14 1423 Rīga Latvia	Tel.: 371-722-5135 Fax: 371-722-4587 sandrao@lbi.lnb.lv
Andrzej Padziński	Main Library University of Agriculture Ul. I Radziszewskiego 11 Lublin Poland	Tel.: 48-81-445-6229 Fax 48-81-445-6229 opracar@eos.umcs.lublin.pl
Anna Paluszkiewicz	Deceased	
Joanna Paształeniec- Jarzyńska	Biblioteka Narodowa Al. Niepodległości 213 00-973 Warsaw Poland	Tel.: 48-22-608-2250 Fax: 48-22-825-8988 bnjparz@bn.org.pl
Edita Peruchoviene	National Library Gedimino pr. 51 2600 Vilnius Lithuania	Tel.: 3705-249-6483 Fax: 3705-249-6129 edita@lnb.lt
Gatis Pogulis	Latvia Education Information System Raina Bulv. 29 1459 Riga Latvia	Tel.: 371-659-5578 gatisp@di.lv
Velta Pozņaka	Research Medical Library Šarlotes ielā 1b 1001 Riga Latvia	Tel.:371-737-8321 Fax: 371-737-3646 velta.poznaka@mzb.lv
Igor Prokop	SNK Nám. J. C. Hronského 1 036 01 Martin Slovakia	Tel.: 421-43-413-1371 Fax: 421-43-422-4900 prokop@snk.sk

<b>Name</b>	<b>Address</b>	<b>Tel./fax/e-mail</b>
Imants Pujāts	Library Information Network Consortium 75 Tērbatas ielā 1001 Riga Latvia	Tel.: 371-708-9019 Fax: 371-708-9489 imants.pujats@rtu.lv
Richard E. Quandt	The Andrew W. Mellon Foundation 282 Alexander Road Princeton, NJ 08540 USA	Tel.: 609-924-3933 Fax: 609-924-8394 metrics@quandt.com
Norma Read	University of Cape Town Private Bag Rondebosch 7701 South Africa	Tel.: 27-21-652-3110 Fax: 27-21-689-7568 norma@uctlib.uct.ac.za
Mihkel Reial	Eesti Rahvusraamatukogu Tõnismägi 2 0100 Tallinn Estonia	Tel.: 372-630-7377 Fax: 372-631-1410 mihkel@nlib.ee
Hans Rütimann	312 West 77th Street Apt. G New York, N.Y. 10024 USA	Tel.: 212- 721-5173 Fax: 212-721-5173 hrutimann@mindspring.com
Krystyna Sanetra	Jagiellonian Library Al. Mickiewicza 22 30-059 Kraków Poland	Tel.: 48-12-423-3786 Fax: 48-12-633-0903 ujsanetr@if.uj.edu.pl
Lýdia Sedláčková	Univerzitná knižnica Michalská 1 814 17 Bratislava Slovakia	Tel.: 421-7-5443-1151 lyda@ulib.sk

<b>Name</b>	<b>Address</b>	<b>Tel./fax/e-mail</b>
Dana Šimková	Computer Science Center Library Department Charles University Ovocny trh 5 116 36 Prague Czech Republic	Tel.: 420-2-2449-1238 Fax: 420-2-2449-1588 dana@cuni.cz
Judit Skaliczki	Nemzeti Kulturális Örökség Minisztériuma Wesselényi u. 20-22 1077 Budapest Hungary	Tel.: 36-1-484-7283 judit.skaliczki@nkom.gov.hu
Maria Śliwińska	ICIMSS Copernicus University Ul. Gagarina 13a 87-100 Toruń Poland	Tel.: 48-56-611-4389 Fax: 48-56-611-4685 maria@icimss.uni.torun.pl
Vija Sproģe	Patent and Technology Library Šķuņu iela 17 1974 Riga Latvia	Tel.: 371-722-6974 Fax: 371-721-0767 vija.sproge@patbib.lv
Jela Steinerova	Comenius University KKIV FFUK Gondova 2 818 01 Bratislava Slovakia	Tel.: 421-2-5924-4562 Fax: 421-2-5296-6016 steinerova@fphil.uniba.sk
Bohdana Stoklasová	Národní knihovna Klementinum 190 110 01 Prague 1 Czech Republic	Tel.: 420-2-2166-3293 Fax: 420-2-2166-3301 bohdana.stoklasova@nkp.cz
Martin Svoboda	Státní technická knihovna Mariánské nám. 5 110 01 Prague Czech Republic	Tel.: 420-2-2166-3402 Fax: 420-2-2422-9224 m.svoboda@stk.cz

<b>Name</b>	<b>Address</b>	<b>Tel./fax/e-mail</b>
Asko Tamme	Tartu Public Library Kompanii 3/5 51004 Tartu Estonia	Tel.: 372-736-1371 asko@utlib.ee
Ágnes Téglási	Oktatási Minisztérium Szalay u.10-14 1054 Budapest Hungary	Tel.:36-1-473-7450 agnes.teglasi@om.hu
Vilija Tolusiene	National Library Gedimino pr. 51 2600 Vilnius Lithuania	Tel.: 3705-249-6483 Fax: 3705-249-6129 vilija@lnb.lt
Dawn Tomassi	JSTOR 120 Fifth Avenue New York, NY 10011 USA	Tel.: 212-229-3700 dt@jstor.org
Signe Tõnisson	Eesti Rahvusraamatukogu Tõnismägi 2 15189 Tallinn Estonia	Tel.: 372-630-7540 Fax: 372-631-1410 signe.tonisson@nlib.ee
Ülo Treikelder	University of Tartu W. Struve 1 50091 Tartu Estonia	Tel.: 372-737-5787 Fax: 372-737-5701 ylo@utlib.ee
Tibor Trgíňa	Univerzitná knižnica v Bratislave Michalská 1 814 17 Bratislava Slovakia	Tel.: 421-7-544-33247 Fax: 421-7-544-34246 trgina@ulib.sk
Ilona Trtiková	CESNET z.s.p.o. Zikova 4 160 00 Prague 6 Czech Republic	Tel.: 420-2-2435-2916 Fax: 420-2-2435-5286 ilona.trtikova@cesnet.cz

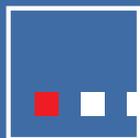
Name	Address	Tel./fax/e-mail
Erik I. Vajda	Magyar Országos Közös Katalogus Egyesület OMIKK cimen Múzeum u.17 1088 Budapest Hungary	Tel.: 36-1-441-4662 Cell: 36-06-30-387-9690 vajdaei@ogyk.hu
Tiiu Valm	Eesti Rahvusraamatukogu Tõnismägi 2 15189 Tallinn Estonia	Tel.: 372-630-7600 Fax: 372-631-1410 tiiu.valm@nlib.ee
Andris Vilks	National Library of Latvia K. Barona ielā 14 1423 Riga Latvia	Tel.: 371-722-5135 Fax: 371-722-4587 andrisv@lbi.lnb.lv
Márta Virágos	University Library Debrecen University PO Box 39 4010 Debrecen Hungary	Tel.: 36-52-413847 Fax: 36-52-410443 marta@lib.dote.hu
Hartmut Walravens	German Union Catalog of Serials Staatsbibliothek 10772 Berlin Germany	hartmut.walravens@sbb.spk-berlin.de
Donald J. Waters	The Andrew W. Mellon Foundation 140 East 62nd Street New York, NY 10021 USA	Tel.: 212-838-8400 Fax: 212-888-4172 djw@mellon.org
Tomasz Wolniewicz	Uczelniane Centrum Informatyczne Uniwersytet Mikołaja Kopernika pl. Rapackiego 1 Toruń Poland	Tel.: 48-56-611-2750 Fax: 48-56-622-1850 Tomasz.Wolniewicz@uni.torun.pl

<b>Name</b>	<b>Address</b>	<b>Tel./fax/e-mail</b>
Graham Woodruff	Innovative Interfaces, Ltd. 3 York Court Upper York St. Bristol, BS2 8QF Great Britain	Tel.: 44-1672-564584 Mobile: 44-772-053-7336 graham.woodruff@iii-europe.com
Petr Žabička	Moravská zemská knihovna Kounicova 65a 601 87 Brno Czech Republic	Tel.: 420-5-4164-6115 Fax: 420-5-4164-6100 zabak@mzk.cz
Éva Zalai-Kovács	Szent István University Villányi u. 29-43 1118 Budapest Hungary	Tel.: 36-1-372-6300 Fax: 36-1-372-6334 ekovacs@omega.kee.hu



The Andrew W. Mellon Foundation and the National Library of Estonia organized a Conference on Union Catalogs which took place in Tallinn, in the National Library of Estonia on October 17-19, 2002. The Conference presented and discussed analytical papers dealing with various aspects of designing and implementing union catalogs and shared cataloging systems as revealed through the experiences of East European, Baltic and South African research libraries. Here you can find the texts of the conference papers and the list of participants.

ISBN 3-937816-08-9



**G A P**

German Academic Publishers