

# Aspects of the Digital Library

Kari Garnes,  
Ane Landøy, Angela Repanovici (eds)

*Alheim & Eide*

AKADEMISK FORLAG

---

*Innehaber Øyvind Nicolaysen*

LAKSEVÅG 2006

Kari Garnes,  
Ane Landøy, Angela Repanovici (eds)

Set in Adobe Garamond pro 12/13,5 pt

(The front page photo by: Pedro Vásquez, 2005)

Cover and inside graphic design by:  
Alvheim & Eide Akademisk Forlag  
innhaver Øyvind Nicolaysen

ISBN 82-90359-76-4

## Table of Contents

|  |     |
|--|-----|
| Foreword .....   | 6   |
| Kari Garnes: About the digital Library at University of Bergen Library .....   | 7   |
| Bjørn Arvid Bagge: Modern technology – modern library premises .....   | 15  |
| Anne Bjørkum Åsmul: The card conversion project at the University<br>of Bergen Library .....   | 31  |
| Halvor Kongshavn & Sverre Sivertsen: The Medical and Law Libraries –<br>and the Digital Challenge .....  | 45  |
| Anne Sissel Vedvik Tonning: The Learning Centre Model at the<br>University of Bergen Library .....   | 57  |
| Maria-Carme Torras & Therese Skagen: User Education at the Digital Library:<br>Physical and Intellectual Access to Information through Digital Literacy .... | 75  |
| Pål Hermod Bakka: <i>Bibliotekportalen</i> - the Library in Cyberspace .....   | 87  |
| Richard David Jones: Institutional Repositories .....  | 111 |
| Rune Kyrkjebø: Parchment and paper in digital University Libraries –<br>new contexts for manuscript and archival collections .....                           | 127 |
| Solveig Greve: Digitisation as a strategy for preservation and dissemination<br>of photographic archives .....   | 137 |
| Susanne Mikki & Elin Stangeland: E-books and their future in<br>academic libraries .....   | 153 |
| Ane Landøy: Implementing the Digital Library – some theories and experiences<br>on leadership of change .....  | 169 |
| Angela Repanovici: Exploitation of Informational Resources within Knowledge<br>Society: Digital Library  |     |
| I. Knowledge society .....   | 183 |
| II. Role of the Library - Infodocumentary Institution within<br>Knowledge Society .....  | 196 |
| III. Exploiting the Digital Library Resources in the<br>Knowledge Society .....  | 200 |
| IV. Data Mining Applied in Automatized Library .....   | 226 |
| V. Text mining applied in digital library .....  | 246 |
| VI. Scientific Research and University Library .....   | 257 |
| About the authors .....  | 283 |

## Foreword

University libraries play a major role in the progress of society, in forming specialists, in promoting innovation and in enhancing competencies. Libraries are destined to be pioneers in the new and emerging fields of scientific development. In the interweaving of information and academic development, the interdisciplinary projects have the greatest success. The collaboration is the main attribute to reach high goals.

The book herein was born from the collaboration between two university libraries: University of Bergen Library, Norway and „Transilvania” University Library of Brasov, Romania. The collaboration came into being by means of the programmes *Leonardo-mobilities under the auspices of the European Union*. This collaboration is materialized through the publication of this book in which the authors reveal the challenges that digital library should respond to.

The book is divided in two parts. The first part contains 12 chapters from Bergen, and the last part contains one large chapter, divided in six sub-chapters from Brasov.

The Norwegian contribution include reports of projects that have been undertaken under the heading “The Digital Library” in Bergen, as well as ideas for how to continue. The reports are hands-on and practical.

The Romanian contribution emphasizes the scientific and research-oriented preoccupations of the Romanian team in the field of information resources exploitation in the knowledge society, where data bases exploitation within the library should lead to an efficient collections management and to an improvement of the scientific research in the academic community.

We strongly believe that this collaboration has been and will continue to be mutually advantageous to both institutions and of great use to readers from all countries of the world.

The editors would like to thank all participants for their time and effort. A special thanks goes to Dr. Rune Kyrkjebø, for his dedication to this project.

Dr. Philos Kari Garnes

Director of University of Bergen Library, Norway

Prof. dr. eng. Angela Repanovici

Director of “Transilvania” University Library of Braşov, România

Cand. Philol Ane Landoy

Arts and humanities librarian, University of Bergen Library, Norway

# About the digital library at University of Bergen Library

By Kari Garnes

At the University of Bergen Library (UBL) the digital library has been developed over a period of 25 years. It started slowly, but during the last decade the changes have been more and more rapid, entailing major alterations to routines and services. There has also been a revolution in the amount of available information. In the first part of this book we will explore the changes and challenges that the University of Bergen Library has been facing during the last 15 years regarding the digital library. The changes are basically of a technological nature, but they have led to major alterations to all the main library processes. Consequently the situation has changed radically compared to 25 years ago, both for the library users (academic staff, students *etc.*) and for the library staff.

## Modern Technology (IT) Developing the Digital Library

The term 'digital library' is ambiguous, meaning different things to different groups.

A two-part definition (after Borgman, C. L. 2002) is useful for the Digital Library:

1. *Digital libraries are a set of electronic resources and associated technical capabilities for creating, searching, and using information. In this sense, they are an extension and enhancement of information storage and retrieval systems that manipulate digital data in any medium and exist in distributed networks.*
2. *Digital libraries are constructed/collected and organized by and for a community of users, and their functional capabilities support the informational needs and uses of the community. They are a component of communities in which individuals and groups interact with each other, using data, information, and knowledge resources and systems. In this sense they are an extension, enhancement, and integral part of informational institutions as physical places where resources are selected, collected, organized, preserved, and accessed in support of a user community.*

Implementing the digital library has entailed both technological and cultural changes. As a consequence of the 'digital revolution' the users nowadays make very different expectations and demands of the library from those they used to make. This means that the library, and consequently also the library staff, has had to change its attitude and way of thinking and working to match user demand. The librarians have to support specific activities in specific contexts, such as scholarship, classroom instruction, education in information, and competition in the context of specific uses and users.

In 1980 the board of the University of Bergen decided to join the BIBSYS library system organisation, an integrated electronic library system that has since become the common system for all Norwegian institutions of higher education. At the University of Bergen Library BIBSYS was first used for acquisition and cataloguing new materials. From 1983 all cataloguing has been done electronically using the BIBSYS system. Since then the main card catalogue has been retrospectively converted into the same online catalogue (OPAC). This has resulted in more efficient and labour-saving library processes. In subsequent years the library began to use the BIBSYS electronic periodicals module and electronic loan system. All the main library processes are currently performed in the same BIBSYS electronic library system.

The BIBSYS system is a central core of the digital library. The content of the digital library includes data, metadata that describe various aspects of data, and metadata that consist of links to or relationships with other data or metadata, whether internal or external to the digital library. Through the digital library more and more electronic resources are made available to patrons. The Internet and web technology have accelerated the development and urge for change.

The number of available resources, with widely varying prices, is constantly growing – digital journals, reference works, databases, and other electronic resources in many different forms, shapes, and formats, including both current and back issues. A new library portal was introduced at the UBL in 2005, common to all the Norwegian universities. The portal is the gateway for accessing electronic resources from the library, and therefore an increasing number of computers have been made available for this usage. Many computers are also equipped to make them suitable for students writing essays, searching the internet *etc.* Institutional repositories with open access (OA) have been developed nationally and internationally in recent years. The library has been instrumental in developing and establishing BORA (Bergen Open Research Archive), which is a digital repository of research output from institutions in the Bergen area. BORA contains full-text peer-reviewed journal reprints, theses, dissertations, and other digital research materials. The library has also started digitising its own material

(pictures and special collections of old and rare books and manuscripts).

The changes that have happened at the University of Bergen Library are explored in greater detail in the other chapters of this book.

### **Building for the new technology**

In the chapter ‘Modern technology – modern library premises’, Bjørn Arvid Bagge takes us through the journey of decisions made when the University of Bergen Library building with its closed stacks was rebuilt to house the Arts and Humanities Library. This meant putting the books on open shelves, as well as accommodating modern technological resources, where patrons, especially students, could use the library as their main workplace, and where, as well using the non-electronic books and journals available, computers were available for use in browsing or searching the catalogue, searching electronic reference works, searching full-text electronic journal articles, or writing essays and assignments. The rebuilt library building also needed to contain rooms for group study or library and information literacy teaching.

### **The electronic catalogue**

The next chapter, ‘The card conversion project at the University of Bergen Library’, tells the story of how the University of Bergen Library converted the card catalogue into the electronic catalogue, making it freely available on the internet for all patrons. The electronic catalogue is the pillar upon which the Digital Library is built. The conversion was a huge job, but in the opinion of the author, Anne Bjørkum Åsmul, it turned out to be well worth it in terms of subsequent labour saving. In addition to being the basis for the Digital Library, the electronic catalogue has also been a necessary tool in the efficient planning of the movement of collections between buildings, or out of the library.

### **The digital library applied**

In the University of Bergen Library, the different faculty libraries have implemented the digital library to a different degree. In the chapter ‘The Medical and Law Libraries – and the Digital Challenge, by Halvor Kongshavn and Svenn Sivertssen, we will read about these two libraries that are among the most digitised of our branches. Both libraries have put a great deal of effort into producing user-friendly web pages, and they are also allocating a lot of resources to user education and on developing close connections to the academic staff.

## **The digital library and the learning challenges**

In the chapter 'The Learning Centre Model at the University of Bergen Library, Anne Sissel Vedvik Tonning describes the background to the projects leading up to University of Bergen's learning centre model. Included in the project were several sub-projects. One of them is further described in the following chapter, 'User Education at the Digital Library: Physical and Intellectual Access to Information through Digital Literacy' by Maria-Carme Torras and Therese Skagen. This chapter explains how, in collaboration with three other Scandinavian higher education libraries, we have developed online learning objects for the virtual classroom to help students improve their digital and information literacy. More specifically, the learning objects have been designed to help students with the information search process and their use of information sources in writing.

## **Why a portal?**

One of the challenges for a university library is how to present the digital and electronic research resources in such a way that they are as efficient as possible for patrons. In the chapter called '*Bibliotekportalen* - the Library in Cyberspace', Pål H. Bakka gives a brief description of the challenges and solutions facing the University of Bergen Library when we were considering buying a portal, and later when we bought and implemented it. He also shows us what the *Bibliotekportalen* looks like to patrons, and evaluates the implementation process, with a special focus on the few things that do not yet work properly.

## **Institutional repositories**

Another aspect of electronic research resources is the escalating price. As an answer to this problem, and the more general problem of getting research published, several university libraries have investigated the possibility of presenting research done at their own universities in institutional repositories. These may also be used for presenting and publishing exams and other student material. The chapter 'Institutional Repositories' by Richard Jones introduces us to the history and current situation concerning institutional repositories, and briefly presents Bergen's own BORA.

## **Digitising our own material**

Especially in the Picture Collection, described in Solveig Greve's chapter, 'Digitisation as a strategy for preservation and dissemination of photographic archives', and the Special Collections as described by Rune Kyrkjebø, 'Parchment



and paper in digital University Libraries – new contexts for manuscript and archival collections’, the University of Bergen Library contains material that may be digitised to give unique content to the Digital Library. In both these chapters the authors present projects currently being undertaken at the University of Bergen Library, as well as their thoughts on how and why to choose the material for digitising. They also discuss particular challenges in archival and preservation practices in respect of unique and perishable archival material.

### **E-books:**

While electronic journals and databases have been on the market for some time, electronic books are a relatively new phenomenon. In 2003, the UBL therefore initiated a project for investigating the process of acquiring, presenting and utilizing such resources. In the chapter ‘E-books and their future in academic libraries’, Susanne Mikki and Elin Stangeland present the results of this project and discuss experiences from the period after the conclusion of the project. Finally, they look at developments in the e-book market and discuss how these will affect the procurement and use of e-books at UBL in the future.

### **Leadership and implementing the digital library**

In the chapter called ‘Implementing the Digital Library – some theories and experiences on leadership of change’, Ane Landøy focuses on leadership in change processes, and uses implementing the digital library as an example of a planned change process.

### **Strategy of the University**

Major challenges in this field are: How can we ensure that our faculty scientists and students get the access they need to the world’s scientific, technological, and other literature and source materials; and: How can we help them develop the skills they need to use that literature effectively?

All the digital developments mentioned here are rooted in priorities in the University of Bergen’s strategic plans and research plans. In the University of Bergen Strategic Plan 2005-2010, developing an institutional repository is a priority, as is developing collections of electronic research materials, both based on our own collections and on the collection of databases, electronic journals, *etc*, that we buy and make available to our users. Electronic teaching materials and learning resource centres are also mentioned in the University’s strategic plan. The University Board has decided to continue developing the digital library at the University of Bergen.

## **Major challenges for the library in the digital age**

With the development of modern technology and today's information society, strong commercial and other interests have entered an area which used to 'belong' more or less exclusively to the libraries. The library sector is therefore faced with an unaccustomed form of competition. What is, or should be, the library's role in this situation? The Internet and web technology develop independently from the thoughts or actions of libraries and librarians. The users often start their searches from a general search engine, and not from a library portal. Contextual information is more than records and documents. Many libraries have unique and valuable collections of older materials. These special collections must be digitised to make them available to the general public. A major task for the library must be to help researchers and other user groups to get easy access to a broad range of electronic resources. It is especially important to assist the scientific community in their scholarly communication, by helping to organise effective systems for electronic research registration and publication, such as institutional repositories like BORA and other open access archives (journals *etc.*). To achieve our goals in this field the library must focus on efficiency, *e.g.* by using systematic performance measurement and developing further systems of library quality control.

A major challenge will be: How are we going to make the library staff skilled for work in the library of the digital age? The staff will have to be continually educated, to ensure that they have the skills and qualifications needed for understanding uses, users and flexibility in performing their duties. To a large degree, staff at the University of Bergen Library already possess these competences, but we still need to develop their skills. Specific skills that will need extra efforts to develop will include pedagogy, especially in teaching Information Literacy so that staff will be able to participate in implementing this in the university's taught subjects. In order to fully achieve this, staff also will need to have an academic background. We also need to make sure that computer literacy among the staff is at a sufficiently high level for them to be able to help and guide library users.

University of Bergen Library is a library with old collections. We will always have books, paper and patrons in our buildings.

**Reference:**

Borgman, Christine (2002) Challenges in Building Digital Libraries for the 21<sup>st</sup> Century. *Lecture Notes in Computer Science*, 2555, p 1-13.



# Modern technology – modern library premises

On the renovation and modernisation of the Arts and Humanities Library at the University of Bergen – how the recent progress in technology and concepts formed a modern library

By Bjørn-Arvid Bagge



*Main entrance, 2005. (Photo: Pedro Vásquez, 2005)*

## Introduction

The traditional book library as we know it is gradually disappearing. Modern technologies, like the world wide web and numerous other internet services, in short *the digital age*, is making its full impact. Together with modern technologies and the development of digital services, there is also a change in the concept itself of a modern scientific library. Libraries are no longer merely to deliver information to students and researchers, but to a larger extent than before to act as pedagogical and systematic mediators of information.

In August 2005 the Arts and Humanities Library in Bergen reopened, after the renovation and rehabilitation of the old University Library main building. This chapter aims at showing how the new digital reality, and the changes in

library management which follow from that, have influenced and given direction to our planning and reconstruction process and thereby formed the new Arts and Humanities Library in Bergen.

After starting with some words about what we planned for, I will say something about the history of our library, then proceed to the planning and construction process. I will then describe how our plans changed, before concluding with a section on what we got, our library of 2005.

### **What we planned for**

Our original building, the old University Library, was a closed stack library building completed in 1961. This building was outdated in more than one respect. Today's requirements in both construction and technology left us only with two options – either a brand new building, or a total renovation of the old one. During the 1990s several committees at our library contributed to an analysis of what functions we wished our modern library to have. On the basis of this analysis a large reconstruction and renovation project was planned.

### **Desired functions – aims and visions – 1998**

The analysis of desired functions were completed by November 1998. Architect Mr. Lars Jarle Nore from the company NORMAN A/S was in charge of this work, on behalf of the University of Bergen Department of Management and Investments, and a group of user representatives from the University Library.

#### **Aims**

- An Arts and Humanities Research Library with collocation of all Arts and humanitites related collections.
- A user oriented specialised research library with open shelf collections and a varied line of services, as far as possible self served by the users (cf. Function analysis (Funksjonsalaysen), 1998, p. 6)

#### **Visions**

- A flexible shape adaptable to future needs
- A library that makes easily available every information recourse
- A library with learning centre functions, able to evaluate, handle and mediate information
- A library with teaching functions tha support and supplement faculty teaching \*(cf. Function analysis (Funksjonsanalysen), 1998, p. 6)

Several of our aims and visions emerged from the fact that we were running an old fashioned subject library where both form and content were, in more than one respect out-of-date. Our original building was already named an Arts and Humanities Library, but it was also seen by many as the general University Library. Reasons for this were partly the fact that several of the University Library's central administrative functions still were located in the building and partly the fact that our building served as a general book and journal repository for the University Library.

Below is a short outline of the history of our old library building, together with a description of the conditions that had to be met before we could start reconstruction and renovation.

## History

### The old University Library

Around the time when the Bergen Museum was founded in 1825, there were several collections of books established, old documents, manuscripts and so forth. Those early collections became the core stock of the Library of the Bergen Museum. When the museum was transformed into a university in 1946/47, the collections from the museum became university collections. Therefore the old Museum Library, originating from the 1820s, is the origin of today's University Library Collections (Gatland et al. 1996: 9-13).

The University Library stayed in the 19<sup>th</sup> century premises of the Bergen Museum until 1961. Following more than 20 years of long bureaucratic discussion, a donation of 5 million kroner from one of Bergen's shipowners, Mr. J. L. Mowinckel, made possible the construction of a modern building for the University Library. On 13<sup>th</sup> of September the new library building had its official opening ceremony, in the presence of His Royal Majesty King Olav V.

#### **Some facts on the University Library in 1961**

- Floor space: 6,350 m<sup>2</sup>, 3 floors, 6 floors book tower
- 500 m area with main catalogues, reference books etc. open to the public
- Student reading room with 230 seats
- Entrance area with exhibition spaces
- 400 m<sup>2</sup> inner area with reception and service area
- 5000 m<sup>2</sup> of closed stacks areas with a total capacity of approximately 15,000 shelf meters

## **From main library to Arts and Humanities Library**

The new University Library in Bergen in 1961 was constructed on the model of old, traditional European university libraries. The users of the library had to place and order for every book they wanted to borrow, and library staff had to pick the book from closed stacks and bring them to the service area where the users would collect them.

At the same time when Bergen opened a new, unified University Library, the trend in other academic institutions was to move from the concept of one large library to smaller branch libraries, bringing library services closer to users and faculty. Already, in 1961, several large research libraries in Europe and the USA were offering users open stacks with direct access to the books. During the 1960's and 1970's the development was clearly in the direction of both open stack libraries and splitting up of central libraries into faculty or department libraries.

These developments also reached Bergen. There was already a medical library service at the university hospital, established in 1956. Since 1968 there has been a Faculty Librarian for Medicine. In 1970 the Medical Library moved into its own premises nearby the university hospital. In the following years faculty library units were established for the other university faculties; these libraries were more or less separated organisationally from the old Main Library.

The later university rector Ole Didrik Lærum lead a committee, with participation by among others our present Library director, Kari Garnes, that decided on decentralising as a main strategy for the University library. The committee presented a framework for how the decentralisation process was to proceed (cf. Gatland et al. 1996: 48-49)

The opening of the Arts and Humanities Library in 2005 marks the final completion of this decentralisation process. We can safely say that the process from main library towards faculty libraries has been a long one – almost 50 years (Gatland et al.).

One precondition for the renovation work to start, was the moving of administrative functions and certain support functions from our building in the summer of 2002. Finally the library building was so empty that it became practically possible to start renovation on it (cf. Function analysis 1998: 3).

## **The planning and construction process**

There was a rapid pace to the planning and construction work, and there were many deadlines to meet. The following outlines the processes that were gone through.



## **From analysis of desired functions to pre-project – the planning process - 1998 to 2001**

By summer 2001 the floor area of the Arts and Humanities Library had increased to about 8,700 m<sup>2</sup>. To be able to stack a constantly growing amount of books, the library had taken over the basement of the Arts and Humanities faculty building, even the one time garage of the faculty staff was now converted into library space. The total book repository at our disposal had now reached 33,000 shelf meters, while Arts and Humanities Collections amounted to somewhere between 27,000 and 28,000 meters. All available space at our library was now occupied by books.

In December 2000 the so called sketch project was started. This project developed further the function analysis of 1998 (cf. User committee, 12.12.2000). The User committee consisted of representatives from the University Library, the University Properties Management Department (EIA), the architect, and other representatives who were called on when needed.

At the time when the User committee was initiated, the first architect Lars Jarle Nore of the company Norman A/S had been replaced by Nicolai Alfsen of Lille Frøen A/S. The function analysis from 1998 was now revised and certain important adjustments done.

Already by December 2000 the suggestion was made that “certain areas of the building should be arranged to serve Information Competence learning” (cf. User committee, 12.12.2000). A study trip to Sweden and Denmark in spring 2001 inspired the User committee with many ideas for how to plan for learning centre activities in our new library. The reform of higher education in Norway at the same time prescribed an increase in student writing activities. This was also an important incentive for us to give priority to information resources. The idea of a learning centre, with emphasis on specially adapted areas for digital media and teaching activities, was also an important aim in our planning.

During spring 2001 our function analysis was adjusted, and plans for the logistic operations of the renovation and reconstruction process were made. Together with the architects and the builder (EIA, University Property Management Department), entrepreneurs were now developing the technical solutions for the transformation of the old Main Library building into a modern faculty library. This work resulted in a document called Pre project (Forprosjekt, June 2001).

Numerous important details of the renovated building were defined in this document. The main floors were to be opened up. Plans were made for a large open space of two floors, around which a mezzanine level was planned, with seats for study or relaxation, and book shelves. Not least, the idea of a coffee bar found it's way to the official planning documents.

It is important here to remember that the basic idea and the plan for the 1961 building was to keep the books, everything except reference literature, under lock and kept away from the hands of the public. The building of 1961 was probably one of the last larger library buildings in Europe to be constructed this way. The main contrast with what we have today, is that the public areas were very small. In fact, general public access was restricted to the reference and counter area, and the reading room, all on the main floor of the building as it stood in 1961.

The former reference area, which was previously dominated by large cabinets with card indices and a large volume of older reference books, was to be replaced by a modern space where digital tools of reference would be prominent. Study rooms for students, teaching rooms, and separate areas for computer use were planned. The old large reading room was to be split up to cater for different functions. At the same time, reading desks and computer desks were planned as an integrated part of library, by being placed in all parts of the building in a spread pattern.

The work leading to the Pre project 2001 was in reality an adaptation of the function analysis of 1998 to our modern technical, or rather digital, world. But the planning process for the transformation of the building did not end there. It was rather that the first step had been taken. Through the whole construction process new decisions were made, and old ones altered because of the rapid development of digital media and electronic technology.

### **The renovation and reconstruction process – moving and building - 2001 to 2005**

As mentioned, the plans for the reconstruction and the related logistics were made in spring 2001. This planning was relatively complicated, because the collections of the library were so large and diverse, and because we had decided to keep our services up and running with only a very short down time planned. Our main questions now were:

- How can we handle our book collections during reconstruction?
- How can we keep running our ordinary library functions, and how do we organise our staff?

A starting point in planning the logistics was to divide the reconstruction work into different phases. An external book repository was established where literature could be kept during reconstruction work on our building. Then the phases

of reconstruction work, the moving of book stocks, and the relative timing of everything, were put down in a precise and detail led puzzle – our logistical plan (cf. Notes and documents, the User committee, March/April 2001).

Some facts are needed to give an idea of the moving and construction process:

- The library staff (around 40 persons), with equipment, was moved; twice.
- A total of more than 9,000 m<sup>2</sup> of office and library areas were moved out of, emptied, reconstructed and fitted up again, and moved into.
- A total volume of approximately 55,000 meters of books and journals were moved.

Moving and reconstruction plans were divided into numerous limited jobs that all had to fit into a greater logistical picture.

Our library stayed open and operational for all but two weeks, during the whole process. Our staff conducted their normal work, only in other areas and under changed conditions. We managed to keep our level of service at nearly the ordinary high level through the whole construction period.

## **New surroundings – change of plans**

### **The automated library**

The book is still today the most important information medium in an Arts and Humanities Library, and it will remain so far into the future. Lack of access to the books was the main problem with the old building. The function analysis stated that one of the chief aims of the new library is that it becomes “a user oriented specialised subject library with books on open stacks” (Function analysis 1998: 6). The library we got, presents itself as an open library where the traditional book shelf is a prominent trait. But even if the book shelves are still there, everything is not as before.

In the new library we have chosen modern, technological solutions for the handling of books. The lending profile of our library was stated in our plans at an early point as “a differentiated service, as far as possible self served by the user” (Function analysis 1998: 6).



*Self served book borrowing.*

*(Photo: Pedro Vásquez, 2005)*

We then searched for an efficient library system that would provide us with a maximum of automated, user self served book handling. Our choice was to install a so RFID (radio frequency identification) system. This system today handles most operations related to the day to day flow of books through the library. Every book is fitted with a radio chip; the system provides a theft alarm function, facilitates self served borrowing of books, self served handling of returned books, automated sorting of returned books, and allows for misplaced books on the shelves to be located.

In connection with the planning and implementing of the RFID system in the library, we were forced to make a number of adaptations of our spaces and equipment. Among other things, we had to find space in the library for the book return automat, and we had to decide where to place lending and return machines and try to fit everything as good as possible with the rest of the equipment and furniture.

The modern building, being an open library, offered considerably more space than the old one. This space we had to fit up with equipment and furniture, and we did so in quite another, more spread and spacy pattern than before.

Implementing the automated library also had some consequences which we at first did not quite comprehend. An example is the choice of book shelves. The RFID technology would work only if we installed wooden shelves, not metal. All of the shelves in our new library were chosen and installed in such a way that we get optimal use of the radio technology.

We may conclude that both the life of the books, and the material and

placing of the shelves and furniture, of a modern library are determined by modern technology, by what could be called, in a broad sense, the new digital reality.



*Book return automat.  
Left: The front, where books are returned.*



*Right: Behind the wall. Transportation line  
with automatic sorting of books.*

*(Photos: Pedro Vásquez, 2005)*

### **Development of the electronic catalogue**

Our oldest catalogues were handwritten card indices. In 1964 photocopiers brought in to duplicate cards. By 1976 we had semi automatic IBM typewriters with electronic memory. Some of us still remember old colleagues back then talking with reverence about the electric typewriter with correction key. It represented a great practical improvement. At the time when the transition to automated cataloguing started in 1983, our card indices had a volume of several cubic meters and occupied a large part of the public area (cf. Gatland et al. 1996: 58).

Computers were first in use in our library from the mid-1970's. The scientific board of our University decided in 1980 that the University Library were to use the electronic system Bibsys as its main cataloguing tool. Cataloguing on paper cards ended in 1983, and the paper card indices were converted to electronic Bibsys entries in the period 1992 to 2004 (cf. Anne Åsmul this volume). Our users can today search our complete book and journal stock without physically visiting the library.

Earlier, the reference area with card indices and voluminous reference book collections occupied a large part of the central open space of libraries. With the transition from paper cards to electronic catalogue, and with the availability of online reference literature, the physical plan of our central open spaces has changed greatly.

Space has been freed to use for other purposes than catalogue and reference. What we have today in our new library is one area for electronic catalogue access, and one area with large round tables with computers for students to work together or alone. The modern digital reality has therefore brought us a much better, more efficient use of physical space also. Both staff and users welcome this development.



*Public functions in the old reference area.*

*(Photo: Pedro Vásquez, 2005)*

### **Library functions – from book library to learning centre**

The internet made its full impact on modern subject libraries from the mid 1990s. With it came the opportunity to collect electronic information from around the globe. From the turn of the millennium and onwards, an ever growing part of the services of libraries are available online.

From a starting point with electronic catalogues, we have seen the development of full text electronic media. Electronic journals and books, and more recently electronic institutional archives (see Jones, this volume) offering swift publication of research, are in many ways changing the landscape of libraries. The physical look and layout of libraries will also probably change.

A large part of scientific literature will in the future be part of the “paperless” world. In the field of journals we already see that the development away from the paper medium has come far, especially in mathematical sciences and medicine. Most of the journals in those sciences are today available electronically. We took this development into account also when planning our new Arts and Humanities Library. With the increased relative volume of electronic publications in mind, we allowed ourselves to cut down the volume of shelves. We could then designate more space for computer desks and study desks, and we had a better opportunity to shape and adjust the aesthetics of the library spaces. This was a freedom we did

not have 10 years ago. Only time will tell if we have interpreted the digital reality correctly and made the right decisions, or if we have miscalculated.

With the transition to electronic libraries there also comes a change in our conception of what a library is. The learning centre idea is today at the very centre of what our university library wants to be. As a learning centre we aim to offer information resources, and an active physical learning space, such as access to computers and study rooms. We will also teach and give courses on the use of our resources and on the larger field of information competence. A library today is not only a keeper and passive deliverer of information, but an active disseminator of intellectual access, information competence and knowledge.

The changes in our mental conception of a library has implications also on its physical infrastructure. Important changes in conception have matured during the period of our planning and reconstruction process, changes that are reflected in the reconstructed and renovated library building as it now stands. A more active role in university education is taken into account and planned for. Heavy investments in computers and digital equipment are made on the basis of a new digital reality and the expectation of larger pedagogical activity in the library. Study and teaching rooms give the building a flexibility to cater for the diverse activities in a modern library.

Developments in technology have shaped the looks of our new Arts and Humanities Library. As we have seen, modern ways of handling the book through an automated library, the electronic catalogue and online resources, together with an active educational role, is the scope and framework for our new library.



*Teaching room.*

*(Photo: Pedro Vásquez, 2005)*

## The library up and running in 2006 – what did we get?

By 2006 our Arts and Humanities Library presents itself as modern and up to date. Teeming with students and teachers, the library has at once become a pleasant and valuable meeting place for people at the university.

Judging by today's situation, most of the strategic decisions seem to have been correct. Teaching rooms with the latest multi media equipment are in frequent use; study rooms for students are very popular and fully-booked. The self served borrowing, return and alarm system is functioning as planned. A large number of computers are at the disposal of our users, both for internet access and for writing of theses and exams. The new furniture seems also to function as planned.



*Free access  
computer arena.*

*(Photo: Pedro Vásquez, 2005)*

### Some facts on the Arts and Humanities Library in 2006

- Opened august 2005 recontructed, renovated and refitted
- Spaces and areas
  - A total of about 8,700 m<sup>2</sup>
  - Closed areas of about 5,600 m<sup>2</sup>
  - Total shelf meters of closed stacks about 12,000 m<sup>2</sup>
- Public areas
  - 3 main floors and a book tower, all with open stacks
  - 3100 m<sup>2</sup> of stacks, computer areas, 5 study rooms
  - Open colletions of about 6,000 shelf meters
  - 70 public computers
  - 120 study desks
  - Wireless internet network in all open areas



## **Physical library and digital services – choices made – what should have been done differently?**

It is not possible to plan for the unexpected. Many people and institutions had a say in renovation the process, and economics is a hard determining factor in many questions. Economic limitations often force us to choose less than perfect solutions.

Something will always go wrong in a large scale process like our reconstruction and renovation. When it happens, when you can not go by the original plans, you have to be creative on the spot and improvise. Then you must have faith and hope that the end result will be good. We often had to do this in our process, in all stages of it, and really in every aspect of it. Unforeseen challenges occurred in the technical field, in relation to furniture and shelves, in several of the moving jobs, and more.

Although preliminary risk analyses were made, in a planning and execution process like the one we went through one is in many respects taken captive by the technological possibilities and development: The digital reality emerges and progresses fast. When the day comes to implement a plan, the basis on which it was made may have changed and made the plan outdated. In our case, automated book lending is an example of this. The technology we implemented in the end, was not available at the time we laid our original plans in 2001. We were compelled to make a change of plans that had quite large consequences for our library building.

The reform of higher education and it's impact on the library also came as something of a surprise to us. Although we adjusted to this we see today that we probably should have used even more space for student study rooms. Neither did we foresee that the collections of the Department of Musicology (UBBGA) were to be located in our library; we had to handle this late in the process and remade some study rooms to make space for the music collections

Technical installations are highly specialised matters and must be planned and constructed by experts. If those experts don't do their job properly, the library can suffer serious consequences at later stages. Technical and safety aspects require a very close and frequent follow up and dialogue through the planning and construction process.

All in all, if the original planning work is good and thorough, most things will fall nicely in place at the end, whether it is digital solutions or other technical challenges. The best way to succeed with a large project like this, is planning, planning and planning. And add to that a good deal of optimism and generosity.



*South façade. (Photo: Pedro Vásquez, 2005)*

## **Bibliography**

- Bonge, Susanne (1984). *Bergens museums bibliotek, Universitetsbiblioteket i Bergen. Fra instituttbibliotek til universitetsbibliotek. En kronike i ord og bilder om tiden på Gamlebygget*. Bergen: Universitetsbiblioteket.
- Gatland, Jan Olav, red. (1996). *Amor librorum nos unit. Universitetsbiblioteket i Bergen 1825-1996*. Bergen: Universitetsbiblioteket.
- Hekland, Brita og Østbye, Linda (1997). Nye biblioteklokaler: planlegging - bygging - flytting - samlokalisering. *Skrifter fra Riksbibliotekstjenesten*, nr 78. Oslo : Riksbibliotekstjenesten.
- Kungliga Byggnadsstyrelsen, utg. (1974). *Universitetsbibliotek*. KBS rapport 65. Stockholm.
- Kårdal, Kirsten, Magni Melvær, Astrid Sægrov (1990). *Sammenslåing av bibliotek. Konsekvenser og gjennomføring*. Oslo : Riksbibliotekstjenesten.
- Lyloff, E. og S. Plovgaard, red. (1984). *Biblioteksbygning 1984. Planlægning af biblioteklokaler i områder med indtil 30.000 indbyggere*. Ballerup: Bibliotekcentralens forlag.

Metcalf, Keyes D. (1986). *Planning academic and research library buildings*. 2. ed. by P.D. Leighton and D.C. Weber. Chicago : American Library Association.

Universitetsbiblioteket i Bergen (1998). [upublisert]. *Funksjonsanalyse utarbeidet for Universitetsbiblioteket i Bergen*. Ved Arkitekt Lars Jarle Nore, Norman A/S. Bergen: Universitetsbiblioteket.

Universitetsbiblioteket i Bergen (2000-2005). [upublisert]. Brukerutvalget for ombyggingen av Historisk-Filosofisk fakultetsbibliotek. Papirer, notater og eposter knyttet til ombyggingsprosjektet 2000-2005.

Universitetsbiblioteket i Bergen (2001). [upublisert]. *Forprosjekt juni 2001*. Utarbeidet for Universitetsbiblioteket i Bergen. Ved Arkitekt Nicolai Alfsen, Arkitektgruppen Lille frøen A/S. Bergen: Universitetsbiblioteket.



# The card conversion project at the University of Bergen Library

On the conversion of our card catalogue to our electronic catalogue Bibsys, challenges and problems in the conversion process, how the project was organized, and benefits of an electronic catalogue.

By Anne Bjørkum Åsmul

## Introduction

When libraries started using electronic catalogues for their collections, it became necessary to convert older card catalogues to electronic versions. In the Nordic countries and Europe this work really got under way during the 1980s and 1990s (Kaltwasser & Smethurst 1992; Beaumont 1989). In Norway the National Library and Oslo University Library converted the University Library Main Catalogue III (HK III) from 1993-1998 (Kiss 1998), and at the Trondheim University Library parts of older catalogues were also converted. To a lesser degree catalogues were converted in other Norwegian academic libraries.

Some of the benefits of an electronic catalogue are:

- all holdings may be gathered into one catalogue
- the catalogue may be accessed and searched from researchers' and students' work-stations, and also from all over the world
- the search possibilities/search combinations/search entrances are augmented
- links to full-text documents
- inter-library lending is easier
- promotion and thereby increased use of the older literature
- participation in promoting the national inheritance
- a basis for automated loan functions
- a basis for efficient routines when ordering and purchasing literature
- a basis for securing and revising collections, especially useful when moving and reshelving holdings
- flexibility when changing catalogue systems
- the import/export of bibliographic records

This chapter tells the story of the card conversion project at the University of Bergen Library. For the main part the chapter builds on the final report from the project (University of Bergen Library 2003).

The project's goal was to convert all the most important card catalogues in the library to our online library system Bibsys. Bibsys is an integrated electronic catalogue for Norwegian research libraries, and the National Library. The system is based on shared bibliographic information, with local copies connected to the bibliographical entries.

During 1992-2003 the project converted the main card catalogue, some shelf lists and institute catalogues. The main part of the work was to convert the Main Catalogue (HK), and this we did from 1994-2003. This chapter deals with the methods, challenges and problems, tells about cooperation with others and the usefulness of the project.

## **Facts and background**

The Main catalogue has entries from 1892-1983, and contains material acquired from the foundation of Bergen Museum Library in 1825 until University of Bergen Library started cataloguing all new acquisitions in the Bibsys database in 1983. The oldest books in the catalogue are from the 14<sup>th</sup> century. The documents covers all subject areas and many languages and nationalities. The catalogue contains approximately 1,03 million cards. Totally we converted 1,1 million cards, and approximately 475.000 records are registered in the database.

The project has had three stages and approximately 33 man-years of work have been completed. The result: **One** electronic catalogue for the library collections.

## **Goals**

The conversion project started in 1992 with the first goal being to register the library journal shelf list as well as a selection of other shelf lists in Bibsys. From the beginning the project was planned to take four years. As the results turned out to be very useful, the project was prolonged. The new goal was to digitize all the important card catalogues in the library by transferring the bibliographic information to Bibsys.

## **The situation in 1993**

From 1982, UBL participated in the Bibsys co-operation; from 1983 all new acquisitions were catalogued in Bibsys. Bibsys uses Anglo-American Cataloguing rules II (AACRII), Norwegian translation and adjustment, and Bibsys MARC-format, (an adjustment of the MARC-format to Bibsys). Our card catalogue was based on "Cataloguing rules for Norwegian Libraries" in different editions, and with local adjustments and rules.

Bibsys will, in addition to its own base, also give access to databases with MARC-data from the Norwegian National Bibliography and the Library of Congress catalogue. These records are easy to copy into the Bibsys base, and may be used as a basis for registering and cataloguing.

When UBL started cataloguing all new acquisitions in Bibsys in 1983, we did not have the capacity to convert the card catalogue at the same time. The card catalogue was closed, and had to be used in addition to Bibsys. This meant that both staff and patrons needed to use two catalogues when looking for literature. We also often had to check both catalogues in the reference work, and after a while the material in the card catalogue was under-used. When patrons did not find the material in the electronic catalogue, they believed that we did not have it. On many occasions it did exist, but was only registered in the card catalogue. Older literature registered in the card catalogue was mini-registered in Bibsys when loaned, but this was just a few titles compared to our large collections. We needed to gather all records in one catalogue. This meant a lot of work, and would be almost impossible to do along with the daily work, therefore it was made into a project, divided into several different parts.

## **Project planning**

Some preparatory work was done before the project application was sent. Already in 1984 the Catalogue Department at the library developed time predictions for a conversion. We estimated that it took around five minutes to convert one card, and using an estimate of the number of main cards in HK we would therefore need approximately 29 man-years of work to convert the entire HK. These estimates turned out to be quite correct.

The National Library converted University of Oslo Library Main catalogue III in 1993-1998. This catalogue contains records of foreign literature from 1966-1979. Here the estimates showed that converting a card took approximately fourteen minutes. By applying this estimate, we would need approximately 81 man-years of work to convert HK. We were certain that we would have a better production rate than this as the National Library conversion was a pioneer work in Norway, which involved considerable teaching of unskilled labour, and moreover at that period Bibsys was partly unstable. We would also be able to build on this conversion because it placed on the Bibsys database literature that we also had in our collections.

Different methods of conversion were investigated, we considered scanning and OCR-reading the cards as there were projects for scanning catalogue cards both in Denmark and Finland. (*MARC-spiste kort*, 1992). However, we soon concluded that scanning would not be useful as our card material was too varied.

The cards had been formatted in many different ways as the cataloguing rules changed, in some cases the printing was poor, and hand-written cards could pose major problems. There would also be a lot of work after the scanning itself, both with the correction of mistakes and the control of duplicates. Also, we would not be able to take full advantage of the common bibliographical data that Bibsys offers, and the advantage of having the registration available in Bibsys immediately would be lost.

Another alternative was to find the books and use them as a basis for the registration. This would require more time, because the books would have to be taken from the stacks and returned, and there would be the problem of all the books being on loan. The already registered bibliographical information that was collected and saved on the catalogue cards would not have been used, but in many cases we would still have to use the cards in addition to the books in order to find the correct bibliographical information.

Therefore, we decided to convert the bibliographical information on the cards directly into Bibsys from the cards in the main catalogue. As seldom as possible would we fetch the books in the stacks, only when it was absolutely necessary in order to check bibliographical information.

Because the total project was so large, we first applied to the University for funding of part of it; we were granted funding of three positions over four years.

We prioritized the following projects:

- the journal shelf list (approximately 23.000 cards)
- some shelf list catalogues and institute catalogues (approximately 59,000 cards)
- the alphabetic catalogue at the Arts and Humanities Library (approximately 70,000 cards).

The selection principles were that the work should be limited, that it should be useful to as many users as possible, and that we should convert the cards only, without fetching the books.

### **Organizing and implementing**

The Conversion Project was organized as a project for card conversion with its own board, consisting of leaders of the different branch libraries and the library director. A project leader was appointed who also had a seat on the board. The board made the principal decisions on what to convert, and supported the project in problematic decisions.

The project started 1992 with three positions available for four years,



one librarian and two secretarial positions. The librarian was to have the daily leadership of the project. We had a large office, located close to the collections and the card catalogue, with three computer work-stations.

Bibsys designated converted entries with their own status in the catalogue, '*konv*', to separate them from ordinary catalogue entries. This signalled to the library staff that the entry was based on conversion from cards, and thus might not conform to cataloguing rules and bibliographical information. Other libraries were freely able to correct these entries if they contained incomplete or inaccurate information.

The work started with the conversion of the journal shelf list. The librarian acquired an overview of the card material, and made routines and rules for this conversion. The secretaries were taught cataloguing and registration rules and initially the librarian proof-read all the work.

## **Method**

We went through card drawers one by one and registered the main cards. All relevant bibliographical information was submitted to the correct MARC fields. The most important fields to be registered were: Author, co-authors, editor, institution, title, sub-title, place, year, number of pages, series, library code, shelf number, classification, and local notes.

Before registering we did a thorough search in the database to see if the document had already been registered. If so, we only added our own local data; library code, shelf number, classification etc. In addition we searched the Library of Congress database and the Norwegian National bibliographical database. During our project more and more catalogues in other libraries were available on the World Wide Web, and some could be used as a basis for the registration (more about this later), this made verifying bibliographical information easier, and added to the quality of our own work.

Cards that were particularly difficult to convert were marked with coloured clips and checked by the librarian. As the staff became more experienced we stopped proof-reading all entries. Each card drawer held 600-700 cards, of which 300 were main cards. On average we had to go to the stacks to verify the books/journals for 5-8 cards in each drawer. Because of this it was useful to be close to the stacks and the catalogue department of the library. During the project we used all the bibliographies and reference works in the library.

Throughout the project, more and more libraries became part of the Bibsys co-operation, and these also in converted parts of their old collections. This has also been useful to us.

In addition to the work of the conversion project, older literature being

loaned was registered in the database by the circulation desk staff. The average percentage of other libraries having already registered the document in Bibsys were around 40%.

### **The first project – the journal shelf list**

The journal shelf list was a part of HK and contained 23.000 cards with all the holdings of journals and series, both present and discontinued. We used about a year (1992-1993) to convert this material. The catalogue contained cards from all the main collections in the library, and covered the time span 1870-1983. The journal shelf list was prioritized because the information would be very useful to the Journals Acquisitions Department in their work with subscriptions and maintenance of the journal collection. Moreover, Bibsys offered a new journals module in the library system from 1995.

The cataloguing rules have been changed during the years, and the cards reflect this. They have been edited in different ways, and rules for names and institutions have changed. There were also hand-written cards, and/or cards that were practically unreadable, so some of the documents had to be checked at the shelves.

Among the particular challenging parts of converting the journal shelf list, we encountered the problem of names for institutions. Both Bibsys and AACRII had other rules than the ones used in this part of the card catalogue. Present practice for institution names is that the name will be written in the institution's native language, using the form adopted by the institution itself. For example in our catalogue we had sometimes registered English forms for Slavonic institutions, and simplified or shortened institution names. Acronyms were also treated differently. It was important to make sure that all the search potentials were maintained, and to make a sufficient number of references.

Series that contained an institution name in the genitive case in connection with a general title like 'Proceedings' or 'Thesis' were also catalogued according to different rules than those applied today. We had to adjust the registration for these, but still have older entries with 'mistakes'. We also encountered challenges when it came to changes in title. Series often change their title, stop, restart, merge and split into sub-series. It was difficult to find correct information about this, particularly when our collection was incomplete. Some languages, e.g. Finnish and Slavonic, posed challenges. It was also difficult to find information about holdings, or in some cases they might contain several series after the present rules. In addition, most of our cards lacked ISSN numbers.

In this part of the project we had quite a good percentage of hits from entries already registered in Bibsys by other libraries, (close to 40 %), largely because the

National Library had already started their conversion of Oslo University Library HKIII, and that other libraries had also prioritized the conversion of journals and series. We also worked closely with our own journal acquisitions department; they received print-outs from all project registrations, and we used their manual cardex catalogue to check incomplete holdings information, missing issues etc. Institution names were verified in reference works and bibliographies, as well as in the ISDS-base (International serials data systems). Some of the entries were given correct institution names and ISSN numbers later.

The conversion of the journal shelf list went faster than planned and was a success. It became even clearer to the librarians and the patrons how useful it was to have everything in one electronic catalogue. Inter-library lending became much easier, and the Journal Acquisitions Department found the electronic version of the journals shelf list very useful, both with regards to the every day work, and as a preparation to use the other functions in the new Bibsys journal module.

All this strengthened the idea and goal of converting the whole card catalogue.

## **Shelf lists and Departmental Catalogues**

From May 1993 to June 1994 we converted other shelf lists and departmental catalogues.

### **Shelf lists**

The library had some shelf lists in card format on collections which were shelved numerically. These cards were quite new (1976-1983), and thus quite simple to register. The cards were also easily readable, and the cataloguing rules used were quite similar to the present. This was literature in high demand, from the subjects philosophy, psychology, pedagogy, languages and social sciences. It was a good starting point for our staff when it came to cataloguing books since it was quite recently published literature, and the percentage of pre-registered material was around 40%.

In addition the shelf list covering the reference collection and the shelf list covering University of Bergen Masters theses were converted.

### **Departmental catalogues**

Some of the University of Bergen departments had large collections of books located in department libraries. These had their own card catalogues, and were supposed to send copies of the cards to the Main Catalogue in the University of

Bergen Library, something which had been done to a varying degree throughout the years. The collections partly contained older literature from 1900 and onwards, but the main part of these collections consisted of literature from 1940-1983, much of it in high demand. Some of these collections were chosen as possible objects for conversion; most were from the Faculty of Science, and the catalogues in the Departments of Chemistry, Physics, Mathematics and Botany were converted. The local librarians informed us about the cataloguing rules for these catalogues, and instructions for converting were made accordingly.

### **Progress and project expansion**

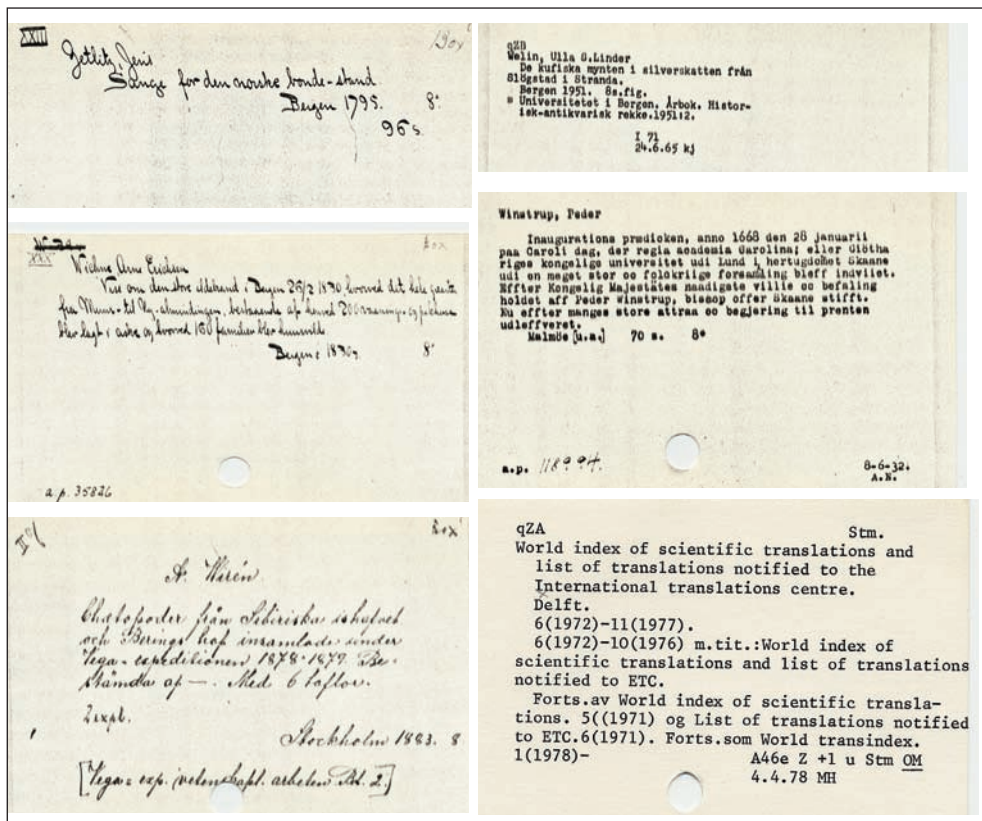
The conversion turned out to be very useful, and in 1994 we applied for an expansion of the project. The project staff was stable, and had acquired experience in reading and understanding the cards, registration, and, when necessary, find books in the collections. The plan now was to start what would eventually become the largest task in the project, the conversion of the Main Catalogue. Work on this started in June 1994, and we estimated that this work would take 10 man-years with the staffing and production that we already had. This turned out to be quite correct, and we finished the work in June 2003.

### **The Main Catalogue**

The Main Catalogue (HK) consists of two parts, one alphabetical and one systematical, the alphabetical part was the one we wanted to convert. Here all the main cards were updated on shelving, number of copies, lost documents etc.

The catalogue contained 1477 catalogue drawers with 600-700 cards in each drawer, a total number of 1,03 million cards, and approximately 700,000 of these were to be converted. The cards had been produced in a 100-year time span, from 1892-1983, and with the oldest literature being from the 14<sup>th</sup> century. We found all varieties of hand-written and typed cards, in different languages, and under several sets of cataloguing rules. From the beginning in 1892 until about 1918 the catalogue cards were hand-written on large slips of paper, and then typed on slips until 1966. In 1967 the slip catalogue was photographed and scaled down to ordinary card format. The scaling down meant that some of the cards were difficult to read, therefore we used the slip catalogue for checking difficult cards.

The cards also varied in quality, but bibliographically the quality was quite good. However documents were shelved in different collections and in different libraries, and the converters had to interpret the locations from the shelf signature on the cards. It was therefore important for the staff to have a good knowledge of how the collections were organized.



Examples of handwritten and typed cards.

## Method

HK was the most demanding of all our conversions. The project librarian had to expand the routines for registration, and add new examples as we encountered new problems. The cards were registered alphabetically according to the main card.

Otherwise the method was the same as earlier; we converted the main cards and did a thorough search in the database before registration. If the record already existed in the database, only local data were added, evident mistakes were corrected and duplicates were merged.

Most of the material has been registered from the data on the cards. Difficult cards (about 8-10 in each drawer) were marked and checked, either at the shelves, on the document itself, or in other sources. Once all the cards in a drawer had been converted, it was marked accordingly, to allow both the project staff and the rest of the library staff to see how far the conversion had proceeded.

### **Quality of registration**

We had to submit the bibliographical information in AACRII and Bibsys MARC-format. Bibsys made a standard for registering entries that did not have catalogue status and this was followed as far as possible. We have incorporated new decisions and rules in Bibsys when applicable. We used Library of Congress and National Bibliographical entries wherever possible, and we used reference works and printed and electronic bibliographies to verify names and other information. As large library catalogues have become accessible on the internet we used them to raise the quality of registration.

### **Problems in the registration**

Since 1991, Bibsys has had an authority file, containing authorized forms of names for people and institutions, and standard titles for anonymous works and series. As one catalogue or registers, data is controlled against this register. One may also put new entries into the authority register, and all references in forms other than those authorized are placed here. In the beginning this control only took place for entries that were catalogued, and we only used this register for checks when we deemed it necessary. From 2001 converted entries were also checked against this register before final registration. This led to better quality for the converted entries, but also made the process more time consuming, as we had to control the correct forms of the name before we could enter them into the authority register. We would register some older literature, being the only Bibsys library to have the document, and we would try as far as possible to enter the correct form of the name according to the present rules. If the author used initials instead of full first name the initials were used. In HK the rule had been to fill in the first name. It could be difficult to tell from the cards what the author had originally done.

Some of the other difficult issues were:

*Author names with different forms:*

Dostojevskij, F.M.; Dostoevsky, Fydor, Dostoejevskij, F.M.; Dostojewskij, Fjodor M; Frans av Assisi, Franciskus av Assisi, Francois d'Assise, Francis of Assisi, etc.

*Royals, peers, and other people with titles or sobriquets* had names of different forms than those used today: Fredrik 2, king of Prussia, Friedrich der Grosse, Friédrick der Grosse etc

*Latinized name forms* in older literature were a problem, also Greek and classical name forms. In our card catalogue we had used classical and Greek name forms, partly adapted to Norwegian. AACRII has a rule saying that in many instances

these should have the Latin form, e.g. Platon in our catalogue was supposed to be Plato in Bibsys.

*Institution names* were to be rendered in the national language; while we often have English versions of the name in HK.

*Literature older than 1800*: Often the information would be sparse and the correct form of personal names might be difficult to find.

Sometimes works were found that were bound together in the same volume, either by the library or by an earlier owner. These could be totally separate works, by separate authors, originally published separately. We would then have to catalogue the separate works individually, and use local notes to explain.

We have also been able to correct shelving errors, and mistakes in the catalogue and in Bibsys, often by communicating with other libraries.

### **Help in the registering work**

The most important help has come from many sources, grouped appropriately below.

The cataloguing rules (Norwegian version of AACRII)

Bibsys MARC

Bibsys rules and decisions

Our own rules and lists of collections and codes etc.

The Bibsys database<sup>1</sup>

The National Bibliography of Norway<sup>2</sup>

The Union Catalogue for Norwegian Libraries<sup>3</sup>

Anonymous classics: a list of uniform headings for European literatures, IFLA 1978

Christensen, Hans Petter: *Greske og latinske navneformer inntil år 1500*, 1995. (Greek and Latin names until year 1500, in Norwegian)

OCLC<sup>4</sup>

Library of Congress authority file<sup>5</sup>

World Cat<sup>6</sup>

---

<sup>1</sup> <http://ask.bibsys.no/ask/action/stdsearch>

<sup>2</sup> [www.nb.no/](http://www.nb.no/)

<sup>3</sup> [www.nb.no/baser/sambok/](http://www.nb.no/baser/sambok/)

<sup>4</sup> [www.oclc.org/](http://www.oclc.org/)

<sup>5</sup> <http://lcauth.dra.com/db/LCAUTH/author.html>

<sup>6</sup> <http://newfirstsearch.oclc.org/WebZ/FSPrefs?entityjsdetect=:javascript=true:screenize=large:sessionid=fsapp2-44394-ekhz14tf-fgv6j1:entitypagenum=1:0>

COPAC (British and Irish Union Catalogue)<sup>7</sup>  
Libris (Swedish Union Catalogue)<sup>8</sup>  
bibliotek.dk (Danish Union Catalogue)<sup>9</sup>  
Karlsruher virtueller Katalog (KVK) (German/Austrian/Swiss Union Catalogue)<sup>10</sup>  
Foreign national bibliographies  
Bibliographic reference works, Norwegian and foreign  
The World of Learning<sup>11</sup>  
Dictionaries

### **Reports and statistics**

We made monthly statistics over the production. The board would meet 3-4 times a year to be updated on the progress. We produced a yearly report and statistics, and we had a large diagrammatic chart in the conversion room where we showed progress by columns.

Number of cards handled 1,119,272

Number of entries converted per year: 44,569

Average entries converted per man-year: 14,856

### **Information and collaboration**

We informed our colleagues on regular basis about the progress of the project. Eventually we had our own web page, where we would update the progress every month, and there was also some information on the library home page. We would give updates at internal seminars and meetings.

Externally we provided information via our web page, on Bibsys meetings, information leaflets and visits from other libraries. We also wrote an article to *Synopsis*, a Norwegian library journal.

### **Advantages of a local conversion project**

There are many offerings in the market for a library to send its card catalogue away and have it converted, but we have seen that a local conversion has proved

---

<sup>7</sup> <http://copac.ac.uk/wzgw?f=f&form=A%2FT&id=6787366>

<sup>8</sup> <http://websok.libris.kb.se/websearch/servlet/se.kb.libris.websearch.Form?type=simple>

<sup>9</sup> <http://bibliotek.dk/>

<sup>10</sup> [www.ubka.uni-karlsruhe.de/kvk.html](http://www.ubka.uni-karlsruhe.de/kvk.html)

<sup>11</sup> [www.worldoflearning.com/views/home.html?authstatuscode=2001](http://www.worldoflearning.com/views/home.html?authstatuscode=2001)



to be very useful. The card material in our catalogue was very varied and it would have been difficult to have this converted without local expertise. Some cards had to be checked against the documents, and this would not been possible with an external conversion. After a while the staff became experts in this, and the quality of the work was very good. The catalogue was immediately available, and mistakes could be rectified immediately. An additional benefit noted by the library staff was that the demand of older literature increased.

The project staff was very stable, and we have not had to train new staff, something that has been a problem for other conversion projects. We also saw that the work was useful almost immediately, particularly when the older literature came into demand, and the rest of the library staff was supportive, all of which encouraged us to continue.

The project personnel have participated in courses in new cataloguing rules, some Bibsys-courses and a short Latin course.

### **For patrons**

All patrons can now access the whole of the University of Bergen Library catalogue online. It may easily be searched from the individual work-stations at the University or from all over the world where there is access to the Internet. The search possibilities in the catalogue have been augmented; one may search further on author and subject, and combined and advanced searches are now possible.

Loans, reservations, returns and messages to patrons have become automated.

The conversion of entries for the older Norwegian literature that few or no other libraries had in their catalogues has contributed to promoting our national cultural heritage. Our national and international inter-library lending co-operation has been improved. We can see that the older literature is still in demand by other libraries, both in Norway and internationally.

### **For library staff**

In the library the staff now has one catalogue, which means that the verification of orders are easier.

Loans, recalls, reservations, returns etc have been automated.

New electronic special catalogues may easily be made.

It is easier to reclassify, reorganize and revise the collections. During the last few years we have undertaken a major reorganisation of collections because

of the rebuilding of the Arts and Humanities Library. This has involved moving collections to open shelves as well as closed stacks. Here we have been able to create lists of collections that were to be moved, change the locality information etc.

We have also started using RFID (Radio frequency identification) for alarm and loan functions. Again, the implementation of the RFIDs has been dependent on the electronic catalogue.

In addition, we believe that the total Norwegian Bibsys database has received an important addition, especially when it comes to older literature. Our newly registered records will function as a basis for other libraries and their future registration, making it be easier for other libraries to convert their older collections when they find our entries in the database.

### **Bibliography**

Beaumont, J., Cox, J.P. (1989) *Retrospective conversion : a practical guide for libraries*. Westport, Meckler.

Kiss, Z. (1998) *Konverteringsprosjektet 1993-1998: sluttrapport*. Oslo, Universitetsbiblioteket i Oslo.

*MARC-spiste kort* (1992). Skrifter fra Riksbibliotekjtenesten, nr 62. Oslo, RBT

Kaltwasser, F.G. & Smethurst, J.M., eds. (1992) *Retrospective cataloguing in Europe*. München, K.G. Saur.

University of Bergen Library (2003) *Universitetsbiblioteket i Bergen: Konverteringsprosjektet 1992-2003: sluttrapport*. Bergen, Universitetsbiblioteket i Bergen.

Universitetsbiblioteket i Bergen. (23.12.03) *Konverteringsprosjektet*, <http://www.ub.uib.no/prosj/Konvertering/index.htm> [accessed February 14, 2006]

Universitetsbiblioteket i Oslo. *Universitetsbibliotekets gamle hovedkatalog - HKI* <http://hki.uio.no/oslo/> [accessed February 14, 2006]

Åsmul, A.B. (1992) Kortkonvertering ved Universitetsbiblioteket i Bergen. *Synopsis*, 2, s. 325-326.

# The Medical and Law Libraries – and the Digital Challenge

By Halvor Kongshavn & Sverre Sivertsen

## Medicine and law

The Law Library and The Medical Library both represent old sciences, medicine with a long tradition at the University of Bergen<sup>1</sup>, law with a rather short one.<sup>2</sup> From a library point of view we share some characteristics, and differ in others. Both libraries represent profession studies, and both faculties have researchers well acquainted with advanced information retrieval systems.

While law may have a focus on existing situation, based on earlier experiences and rereading of law sources, medicine is focused on development and research in new fields. Even if law is developing in new fields, and medicine have strong traditions and need to reread history, we may say the medical information need is characterized by its currency.

Research and teaching are of course interwoven, and the number of students compared to the number of researchers may have an impact on how we organize our library services. The Faculty of Law in Bergen has approximately 2000 students registered, while the Faculty of Medicine has 1300 students.

While the medical library spends most of its budget on journal and database subscriptions, the law library has a relatively greater need for books.

In this chapter we give some general introductory remarks on the two faculty libraries, present information resources important to the libraries and their patrons, describe the libraries as part of the learning and research environment, and present some challenges that we consider important for a successful future development.

## History

The Medical and the Law Libraries were established in the early 70's, the Medical Library with a predecessor on medicine as early as 1956 at Haukeland Hospital.

<sup>1</sup> The study of medicine was introduced in 1946, and the Faculty of Medicine was established in 1948.

<sup>2</sup> The law study started up at the University of Bergen in 1969, from 1970 under the Faculty of Social Science until the Faculty of Law was established in 1980.

The Law Library was until the early 90's formally organised as a part of the Social Science Library. The Law Library moved to new premises in 1995, The Medical Library in 2003.

Medicine and law were, in different ways, pioneers in the development of electronic information, and have a history of sophisticated systems of printed information storage and retrieval, like manuals, bibliographies and indexes.

### **From Index Medicus to PubMed**

The printed reference works for the biomedical literature dates back to the US Library of the Surgeon General's office, and their first printed catalogue from 1864. In 1874 they started indexing the journals in the library, and Index Medicus appeared as a monthly publication in 1879. Index Medicus was a classified subject guide with author index, which made it possible to keep up with the current research results published in medical journals.

In the middle of the 1960s the Index Medicus went electronic with MEDLARS, The Medical Literature Analysis and Retrieval System, the largest publicly available information retrieval system of the time (Blake, 1986, Miles, 1982). In the beginning one had to order searches from specialists, then magnetic tapes were distributed to selected libraries, and later floppy discs and CD-ROMs were offered to be installed on single computers or on servers. The final product was web based versions of MEDLINE, and the most well known version of Medline, PubMed<sup>3</sup>, is now available without costs from The National Library of Medicine.

### **Law**

Within law the electronic age started in the US with "Legal Information Thru Electronics" – LITE – in 1963, and in Europe the first system was CREDOC in Belgium from 1967.<sup>4</sup> The electronic systems were based on the printed ones. In Norway the printed Supreme Court law report was established in 1836<sup>5</sup>, the history of published acts and regulations is older than the history of print. The different collections have always had more or less sophisticated systems of indexing. The predecessor of the legal information system Lovdata<sup>6</sup>

<sup>3</sup> [www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=PubMed](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=PubMed)

<sup>4</sup> From Bing (1989). Professor Jon Bing, University of Oslo, has been – with director Trygve Harvold in Lovdata – one of the core persons in the development of legal information systems in Norway from the very beginning. Both Bing and Harvold are still very central in this work.

<sup>5</sup> "Norsk retstidende" still runs.

<sup>6</sup> Some information in English is given at [www.lovdata.no/info/lawdata.html](http://www.lovdata.no/info/lawdata.html).

started its' work in 1979, and the opening to the public was in 1983. Governmental information, including legal documents, was available online from 1995 in ODIN<sup>7</sup> – Offentlig Dokumentasjon I Norge (Public Information in Norway<sup>8</sup>), while the parliament's web-service<sup>9</sup> was launched in 1996.

Besides these important three, a number of other more or less specialised databases on Norwegian, foreign and international/multinational law have been introduced through the recent years, and the law library subscribes to an increasing number of these. Some are also available for free. The three academic law libraries at the universities of Oslo, Bergen and Tromsø were established in 1956, 1974 and 1987 respectively.<sup>10</sup> We have cooperated closely through the years. In 2000 the law web portal Juridisk nettviser<sup>11</sup> was introduced as a joint product from these libraries, which later also included the Norwegian Bar Association as a partner. The potential in the digitalisation of the primary legal sources was utilised at an early stage, as pointed out. On the other hand – secondary materials like journal articles and books were not presented in digital version until recently. This is partly due to a general conservatism, partly to the fact that law is a “book-subject” and partly a question of copyright. This is now under change, due to the success of the Lovdata online information system and the general development of digital journals combined with satisfactory solutions of copyright questions. Today law students tend to overlook relevant sources of law only available in printed versions.

Law is basically and principally national, and in Norway mainly written in Norwegian for a Norwegian audience. On the other hand; the influence of and relation to international and foreign law is of more significance than ever – reflecting the general globalisation. This has the last decades given new and increasing challenges to the library and its information services, including both acquisition and competence. What the Swedish law professor Peter Wahlgren wrote in 1999, is even truer today:

“The fact that several legal systems are approaching each other implies, among other things, that the volume of legally relevant material is growing drastically. In addition, the rate at which legal material is published increases all the time” (Wahlgren, 1999:38).

---

<sup>7</sup> The ancient Norse god Odin left his eye in the purifying waters of Mimir's spring for wisdom (English Wikipedia March 2006).

<sup>8</sup> The English pages start at [odin.dep.no/odin/english/bn.html](http://odin.dep.no/odin/english/bn.html). The predecessor of ODIN was DOP (Dokumentbase for offentlige publikasjoner = Document Base for Public Publications), originally run by Lovdata.

<sup>9</sup> [stortinget.no](http://stortinget.no)

<sup>10</sup> See Bertnes, P. A. & Kongshavn, H. (2006) for more details on Norwegian law libraries.

<sup>11</sup> [juridisk.net](http://juridisk.net)

## **BIBSYS and the web**

Public access to the online catalogue BIBSYS<sup>12</sup> around 1980 aided researchers within both law and medicine – in particular within law which to a large extent is a “book-subject” in the Scandinavian countries.<sup>13</sup> During the approximately 30 years of BIBSYS, the system has developed a lot and is today an integrated part of everyday life both for library staff and our users. Besides, the national article reference bibliography NORART<sup>14</sup> developed, and is of value for the work in law libraries. Medicine is internationally orientated, and PubMed and other international databases index only a few Norwegian journals. To search for health related publications in Norwegian or other Scandinavian languages, the health community has to rely on NORART and the database Svemed<sup>15</sup> produced at the Karolinska Institutet in Stockholm, Sweden. The University of Bergen Library went on the web in 1995. Both medicine<sup>16</sup> and law<sup>17</sup> have been very active in the developments of their web pages, which has brought them to be the most visited departmental web sites at the library.

The use of the online systems has increased slowly. The start was very limited with one user, modem connection and terminals in the library. Now every student and researcher at the University of Bergen has access to an amazing amount of online information, both references and to a larger extent full text version. The library has changed, but the tasks remain pretty much the same.

## **Where are we now?**

We are definitely at a turning point. The importance of printed documents is decreasing rapidly and more rapidly within medicine than law. Does this mean changes when it comes to the understanding of the concepts of “library” and of “librarian”?

## **The Medical Library**

The Medical Library is representing a wide field of research and teaching, from molecular biology to international health. This means that we must offer a wide

<sup>12</sup> [ask.bibsys.no](http://ask.bibsys.no)

<sup>13</sup> This varies around the world, i.e. in the United States academic law libraries today use up to 90 % of their acquisitions budget on non-book materials.

<sup>14</sup> See English version [www.nb.no/baser/norart/english.html](http://www.nb.no/baser/norart/english.html), references to articles from 390 Norwegian periodicals and yearbooks.

<sup>15</sup> [micr.kib.ki.se/netahtml/arti.html](http://micr.kib.ki.se/netahtml/arti.html)

<sup>16</sup> [www.ub.uib.no/avdeling/med](http://www.ub.uib.no/avdeling/med)

<sup>17</sup> [www.ub.uib.no/avdeling/jur](http://www.ub.uib.no/avdeling/jur)

selection of resources both in the physical library, and on the Web. For the medical researcher PubMed (or Medline in another version) is still the most important database. The number of references in PubMed has now exceeded 16 millions, and every month around 40.000 new references are added. ISI Web of Science<sup>18</sup> is also a database much used to search for research results in biomedicine, and getting access to the articles through the connection to our collections. As a needed supplement to PubMed, researchers who want to publish in scientific journals are often advised to search the more European oriented database EMBASE (Excerpta Medica Database)<sup>19</sup>.

The librarians at the Medical library always strongly suggest that researchers and students use several databases if they need a good overview over what has been published in their field. A lot of research projects also involves more than one faculty (pharmacy, psychiatry), and the field of interest is not covered only in medical databases. We therefore focus on the databases PsycInfo<sup>20</sup> and Biological Abstracts<sup>21</sup> in our presentations on the Web and during courses.

The medical library in Bergen is also the research library of the Haukeland University Hospital. In the clinic the need of information is often more focused towards diagnosis and therapy than research, and we promote the evidence based medicine databases Cochrane Library<sup>22</sup> and Clinical Evidence BMJ<sup>23</sup>. PubMed also has “clinical queries” searching, with predefined search filters to locate clinically oriented publications.

These databases are also of interest among user groups like nurses and midwives, which have used the library and its electronic resources to an extended degree during the last years. Nursing literature may be found in the already mentioned databases, but also in the nursing database CINAHL (Cumulative Index to Nursing and Allied Health)<sup>24</sup>.

Open resources with medical information on the net are mostly indexed in BIBSYS when it comes to publications like books, journals or databases, and in the BIBSYS Subject Portal<sup>25</sup> (BIBSYS Emneportal) if it is web pages dedicated to certain subjects.

How does the library promote its services? The starting point on the web is the subject resource page, where we obviously have links to the library portal and BIBSYS Ask catalogue and other information of interest: How to publish, news

---

<sup>18</sup> [www.isinet.com](http://www.isinet.com)

<sup>19</sup> [www.embase.com](http://www.embase.com)

<sup>20</sup> [www.apa.org/psycinfo/](http://www.apa.org/psycinfo/)

<sup>21</sup> [www.biosis.org](http://www.biosis.org)

<sup>22</sup> [www.cochrane.org](http://www.cochrane.org)

<sup>23</sup> [www.clinicalevidence.com/ceweb/conditions/index.jsp](http://www.clinicalevidence.com/ceweb/conditions/index.jsp)

<sup>24</sup> [www.cinahl.com](http://www.cinahl.com)

<sup>25</sup> [emneportal.bibsys.no/ep/cgi-bin/ep](http://emneportal.bibsys.no/ep/cgi-bin/ep)

service (new databases journals, books and so on), and open courses on how to search different databases and reference handling programs like Endnote.

We promote the library towards the students on different levels, and the library has for many years done courses as an integrated part of the doctoral education programme. The researchers are contacted by visiting institutes and departments with special courses, or introduction to library services during lunch breaks. We also offer a “Book a librarian” service, where researchers and students can get help with literature searching.

A lot of the resources we offer may be accessed without knowing that the library is offering the services (like journal articles in full text on the net). This may mean that the library is not recognised as the most important provider of research sources any longer (Borgman, 2002).

### **The Law Library**

In 2006 Lovdata, Stortinget and ODIN give access to all sorts of primary legal materials for the Norwegian lawyer; that is pre-law – or preparatory – documents starting from the 1990’s; all acts and regulations in force; and reports from the Supreme court from 1945<sup>26</sup> and from the appeal courts from 1993<sup>27</sup>. Stortinget.no and ODIN are freely available, both with loads of easy accessible information on law (and more), while Lovdata has become close to a complete library – of acts, regulations, reports from courts and other law enforcing bodies, full text articles from law journals, EU materials and treaties which include Norway as a part, etc. Lovdata’s free web version includes complete full text consolidated acts (published in 1995) and regulations in force (published 1997).

The web portal Juridisk nettviser provides “user friendly and [...] quick access to legal information and other information which could be useful in research, studies and other practice – all in a Norwegian context”<sup>28</sup>, all together around 1400 registrations in 2006.

The conservatism regarding online publishing of articles is changing. The last couple of years we have seen a radical change with the development of Idunn<sup>29</sup> – full text articles from the main academic journals publishers from 2001, including law – and within Lovdata with more than 2100 full text articles.<sup>30</sup>

---

<sup>26</sup> In addition comes all the “important” reports from the period of 1836-1945.

<sup>27</sup> Reports from one appeal court are available from 1991.

<sup>28</sup> From the English pages of Juridisk nettviser – [juridisk.net/english/info](http://juridisk.net/english/info). Juridisk nettviser has an English interface at [juridisk.net/english](http://juridisk.net/english)

<sup>29</sup> [www.idunn.no](http://www.idunn.no)

<sup>30</sup> 2141 articles by 16 March 2006. The numbers are increasing every month, and each article is equipped by Lovdata with hyperlinks to acts, reports, etc – giving a substantial added value.



The use of the web accessible information has, as pointed out, literary exploded through the recent years, while the use of the analogue library has decreased some. At the same time the use of the library facilities has increased quite a lot, due to the introduction of both the medicine and the law libraries as so called “learning centres” (Tonning, 2006, this volume)

The Law Library is successfully integrated in the faculty system, and has for many years done courses on different levels as part of the curriculum. The courses include legal research, library knowledge, introductions to relevant databases, law information literacy<sup>31</sup>, the use and quoting of the sources of law. The library’s premises are in the centre of the law faculty building, very close to the students who have to pass the library several times every day to and from their activities in the building and the faculty who has a separate entrance to the library premises. We try to maintain a high profile and the library is active in promoting its’ services in various ways, e.g. through the “Tip of the day”<sup>32</sup>, through “Book a librarian”<sup>33</sup>, through our news on the front web site, through our extensive student pages<sup>34</sup>, through exhibitions<sup>35</sup> and through publications, including a textbook in legal research<sup>36</sup>.

The integration with the faculty is very valuable when it comes to choose the acquisition of resources – books, journals, electronic sources. We communicate both formally through the circulation of prospects and through our “Suggest a book” service<sup>37</sup>, and informally through daily conversation, lunches etc.

## Where do we go from here?

The University of Bergen Library is planning to transfer all journal subscriptions from paper to electronic versions during the years 2005-2008. This will have some obvious benefits like full access to electronic journals from all computers in the university and hospital networks.<sup>38</sup> This also means that we no longer have possibilities to store our own archives of journals, and we have to depend on good internet connections, with dependable providers to offer continuous access.

---

<sup>31</sup> The term “law information literacy” is used here for the first time.

<sup>32</sup> See archive from 1998 on [www.ub.uib.no/fag/rettsvit/dt/index.htm](http://www.ub.uib.no/fag/rettsvit/dt/index.htm).

<sup>33</sup> [www.ub.uib.no/avdeling/jur/arkiv/bestillbibleng.htm](http://www.ub.uib.no/avdeling/jur/arkiv/bestillbibleng.htm)

<sup>34</sup> [www.ub.uib.no/fag/rettsvit/jus-rettskilde.htm](http://www.ub.uib.no/fag/rettsvit/jus-rettskilde.htm)

<sup>35</sup> We have several exhibitions every year, i.e. in this Henrik Ibsen year of 2006, which is 100 years after the author’s death, an exhibition of Ibsen and the law will be opened – with particular emphasis on the use of contemporary law in Ibsen’s plays. Each exhibition is followed by a text on the web, check the archive starting in 1996 on [www.ub.uib.no/avdeling/jur/arkiv/utstillinger.htm](http://www.ub.uib.no/avdeling/jur/arkiv/utstillinger.htm)

<sup>36</sup> Bertnes and Kongshavn, 2005.

<sup>37</sup> [www.ub.uib.no/avdeling/jur/arkiv/suggestabook.htm](http://www.ub.uib.no/avdeling/jur/arkiv/suggestabook.htm)

<sup>38</sup> And – of course – the premises vacated will provide more workspace for students and other users.

There will no longer be increasing demands for storing space, and we may decide to get rid of paper versions already electronically accessible. What do we do next? We may let the National Library be the sole owner of paper versions already paid for and catalogued. This will secure access to old volumes, if something should destabilize the electronic delivery system.

### **Paper out, computers in**

Less need for storing paper means more space for other activities. The medical and law libraries have already started to remove shelves and are replacing them with computer working places for students. This trend will continue. There may also be need for teaching rooms, like computer labs in the library area, since the demand to promote and teach the library resources will increase.

The medical researchers approve totally to the expanded electronic access policy, and we have experienced only few negative reactions to a complete stop in paper subscriptions. When the Medical Library started with electronic subscriptions and – among other changes in our routines – ceased circulating paper copies some patrons were opposing. The paper copy was considered the real journal, and the electronic version an extra service. Today the attitude is totally opposite, and we experience that users now tend to reject paper copies, and rather pay for electronic access with their own credit cards. This change in attitude has evolved in only a few years time, and may also be seen among the users of the law library.

We may choose to see the reduction in paper subscriptions in a more humorous way, and be thankful for not having a library as described in Terry Pratchett's Discworld books: "Wise students in search for more distant volumes take care to leave chalk marks on the shelves, and tell friends to come looking for them if they're not back for supper. Even wiser students don't go in at all" (Pratchett and Briggs, 2003).

### **Finance**

One of the challenges in collection development is now to be able to finance access to the most important databases, journals and other online resources. The subscription prices have been rising much faster than the general increase in prices, and value added tax (25%) on electronic journals in Norway, in contrast to paper journals that have no such tax, makes the budgeting even more difficult. An example illustrates the challenges in making electronic access to scientific journals: Although the University of Bergen Library now offer access to around

13.000 journals online, the Medical Library had to order articles from more than 1000 different journals in 2005.

Also; to secure access to older journal volumes we should give a high priority to buy good archives.

### **Something lost, much gained**

Have we lost something on the way? For the researchers in medicine the access to journals from their own computer, have led to a nearly zero contact with the physical library, tendencies we also can see within law. And at least in the Medical Library this has obviously led to less direct contact with the researchers. These days the contact is mostly through suggestions of new subscriptions or new books, or complaints when the data network is not functioning. This leads to a greater demand for the library to contact the researcher, or the researchers' representatives at the faculty, to discuss library topics.

What have we gained? Besides the positive aspect of increasing electronic access, we have become much more a library for students (meaning the actual library room). The Law Library has always focused on accessing student material, while the Medical Library without doubt has had a priority towards the researchers. The new study reforms at the university with stronger follow-up from the teachers, more writing, problem based learning and more have forced us to have a stronger priority towards the students. This has manifested itself in more work places and computers in the library, and a focus on the need of student literature. Even in a journal based science like medicine students still need a good book collection.

### **New possibilities**

The libraries' websites present resources for studies and research, and we offer ways of getting in touch through phone or e-mail. These are pretty standard ways of communication. May we use some new technology to strengthen the contact, including again getting close to the researcher? Library workers have traditionally had a rather formal and regulation oriented attitude. To assure quality is a strong value and this has led to good catalogues and good information retrieval systems. Should we still loosen up a bit, e.g. by setting up wikis<sup>39</sup> on the library home page? Is it possible to have chat groups for students, researchers and library people?

One of the more popular new ways of using the web is by web-logging (blogging). Private stories, political writings and hobby topics are easily presented

---

<sup>39</sup> The encyclopaedia Wikipedia is an example of new ways of using the web.

on the web in a web-log. Each librarian may perhaps have his own blog to share new ideas or raise discussions related to the library. This is a system that opens for easy communication, since comments may be posted directly to the last publishing on the blog, and the library will be dynamic and interactive (Plutchak, 2005).

The medical and law libraries give a large number of courses and presentations every year. Could oral presentations or perhaps videos be presented on the website? With the technology oriented student it will perhaps be just as natural to listen to a PodCast or watch a presentation with sound and pictures in a videocast on the screen? Utilizing these technologies may increase the interest of the user towards the library services. The technology for setting up blogs and podcasts are inexpensive and easy to use, and therefore adapted by a lot of users in their leisure time (Cochrane, 2005). We also see that scientific journals like Nature are presenting podcasts.

How about the library staff? Do we need librarians with special skills for cataloguing, or will most of the resources we access be supported by additional material like keywords and signatures ready to put on the shelf or in the network? Do we need staff with degrees in the topics we represent in the library, to support the need of the student and researcher? We still need staff to fetch journals or books to copy or lend out. We still need librarians to organize the information, and people with degrees or a very good understanding of the field researched or taught at the faculty. We also have to be more and more like teachers, teaching how to use the different sources we supply, either in the library, on the campus visiting departments, or as part of the teaching program, where searching for information is a natural basis for the work students have to do. But still – the most useful and important contact with our users is in the one-to-one basis.

We do not think that the student and researcher will be fully satisfied with search robots in the near future, even if they get better and may help the internet surfer in many ways. An editorial in BMJ even states that building “Google medicine” would be of immeasurable benefit to human health (Giustini, 2005). The information should be instantly accessible from anywhere the student and researcher are working, and the library is the logical entity to manage the resources needed (Lindberg and Humphreys, 2005)

## **New roles?**

The decentralised faculty libraries of medicine and law are successful, in particular when it comes to the focus on and the closeness to the main users – the faculty and the students. This includes the necessary flexibility when changes are needed.

A lot of important and relevant challenges happen around us and at University of Bergen Library and in our local faculty libraries.

But the basics of library thinking remains – the British library professor Peter Brophy concludes in 2001 in his book “The library in the twenty first century: new services for the information age” thus:

“There are many, highly valued, roles that libraries can fill in the 21st century, but the most powerful will be the ‘community information intermediary’, a body that understands and has empathy with its community of users, has deep understanding of the information universe and its organization, and actively develops and promotes the mechanisms that link the two together” (Brophy, 2001:184).

We agree. The need for libraries and library people has never been higher. We have a bright future.

### **Bibliography**

- Bertnes, P. A. & Kongshavn, H. (2005) *Praktisk rettskildelære: juridisk informasjonsøking*, Oslo, Gyldendal akademisk
- Bertnes, P. A. & Kongshavn, H. (2006) [On Norwegian Law Libraries]. *Scandinavian studies in law*, 49 [in print]
- Blake, J. B. (1986) From Surgeon General’s bookshelf to National Library of Medicine: a brief history. *Bull Med Libr Assoc*, 74, 318-24
- Borgman, C.L. (2002) Challenges in building digital libraries for the 21<sup>st</sup> century. *Lecture notes in computer science*, 255, 1-13
- Brophy, P. (2001) *The library in the twenty-first century: new services for the information age*, London, Library Association Publishing
- Cochrane, T. (2005) *Podcasting: the do-it-yourself guide*, Indianapolis, Wiley
- Giustini, D. (2005) How Google is changing medicine. *Bmj*, 331, 1487-8
- Lindberg, D. A. & Humphreys, B. L. (2005) 2015: the future of medical libraries. *N Engl J Med*, 352, 1067-70
- Miles, W. D. (1982) *A history of The National Library of Medicine: the nation’s treasury of medical knowledge*, Washington, National Library of Medicine

Plutchak, T. S. (2005) I see blog people. *J Med Libr Assoc*, 93, 305-7

Pratchett, T. & Briggs, S. (2003) *The new Discworld companion: the fully revised and updated reference book to all things discworldian*, London, Gollancz

Wahlgren, P. (1999) *The quest of law: law libraries and legal information management of the future*, Stockholm, Jure AB

# The Learning Centre Model at the University of Bergen Library

By Anne Sissel Vedvik Tonning

## Introduction

At the start of the new millennium, general access to a great variety of sources of information was increasing exponentially, and a concern that students acquire skills enabling them to assess the quality of these sources was being voiced by many. Another, related challenge was the growing concern of plagiarism in students' written work.

At the University of Bergen Library (UBL) the time was thus ripe for change; a new set of student services were needed. Until this time, the principal focus of the library had been the information needs of university researchers. For instance, students did not have access to all library services until they had reached the master level. However, after an official national report of higher education (Ministry of Education and Research, 2000), it became clear that our focus would in future need to include the student body in its entirety. The report described the institutional libraries and their services as vital resources in student learning – for students on all degree levels.

The report led to the Quality Reform of Norwegian higher education, accepted by parliament as Stortingsmelding 27 (Ministry of Education and Research, 2001). This reform, the Norwegian reply to the Bologna process, was to be implemented by the start of the academic year 2003 at all institutions of higher education in Norway. Central emphasises of the reform were: developing methods of learning and teaching geared towards active student participation; information and communication technology (ICT); closer supervision of students; new forms of examination and assessment; quality assurances; a new degree structure; internationalisation; and a reorganisation of the system of students' public support. Also, part of the basis for the allocation of funds to each institution would no longer be its number of registered students, but the amount of course credits accumulated by those students.

These changes affected the UBL and its services in several ways. Firstly, the institutions were required to provide a student library with services that would form an integrated part of the educational programme. Secondly, the new emphasis on active student participation in learning and teaching methods would also require adjustments. For example, by way of teaching students to

assume responsibility for their own learning process, the old set reading list would largely be replaced by self-chosen texts based on individually formulated discussion topics. Inevitably, this would entail an increased need for library services. The UBL could thus legitimise its need for change and development in terms of the Quality Reform. The library would now cooperate closely with academic staff to develop suitable learning environments for the students, and to improve their information literacy. The library could potentially become a vital pedagogical resource, contributing students in finishing their courses on time, and thus securing funding for the faculty based on the course credits produced.

So, how would the UBL in practice improve its independent student learning resources in accordance with the demands of the Quality Reform? In 2000, a group of library administrators visited the Sheffield Hallam University learning centre; for a number of us, this inspiring visit made the thought of establishing learning centres of our own an immediately attractive one. But what would our own version need to look like in concrete terms?

After having outlined the concepts of learning centre and of information literacy, the remaining of this chapter will offer a retrospective view of the process – the projects and activities undertaken – towards the development of the UBL learning centre model. The projects involved were initiated and carried out by UBL staff as bottom-up projects; they were founded on strategic documents and public reviews, and supported by the library management. Furthermore, the chapter will provide an assessment of the kinds of challenges organisational change may entail, with specific reference to the implementation of learning centre at UBL.

### **Learning Centre**

In Europe, the term learning centre came to prominence within the higher education sector during the 1980s re-organization of this sector in Great Britain. As a consequence of the reform which turned former polytechnics into universities, as well as of rapid technological change in society at large, a clear need arose to expand students' library and learning support resources at universities nationwide. The 1990s saw an extensive co-ordination of supporting services for teachers, students and faculty members in institutional libraries; this was called a learning centre (Abson, 2003). This initiative started a new trend in library development, soon to be adopted in the Nordic countries and other European countries as well. In Norway, too, higher education libraries started to investigate different learning centre models. UBL's definition of a learning centre is based on the definition of Fagerli (2000), which focuses on the physical provision of facilities within a pedagogical framework. It emphasises the following factors:



- Library, access to workstations for individuals as well as for groups, etc.
- User support
- Pedagogical adaptation
- Cooperation between teaching staff and supportive services

From the beginning, the UBL intended to place significant emphasis on the real content of the learning centre. We did not want the learning centre to end up functioning simply as another computer room. However, this of course raised questions: which services would we prioritise, and what kind of pedagogical framework did we envisage? An important concept in this discussion was that of information literacy. This concept brought those aspects of the learning centre model especially relevant for the UBL – sifting the vast flow of information world-wide, and the problem of plagiarism threatening academic integrity – into sharp focus.

### **Information Literacy**

No common standard of information literacy has been established either nationally or on the level of inter-Nordic cooperation. The information literacy projects at UBL therefore rely on American (American Library Association, 2000) and Australian (Council of Australian University Librarians, 2001) information literacy standards, in particular the following aspects:

- Recognising a need for information and defining the problem
- Choosing and searching sources of information
- Source criticism
- Using and quoting relevant sources
- Applying information in written work

These points also clarify the relationship between the student, the library, and the academic teacher (Bruce, 1997). The student is responsible for his or her learning process, and accordingly needs to recognize a need for information. The information specialists in the library are the specialists in information sources, gathering and retrieving information, including source criticism, the use of sources regarding reference techniques and ethics. The teachers are the specialists in supervising the student for the use he or she makes of information in his or her own texts.

These factors are also central to developing the kind of competence defined by the Programme for Digital Literacy (Ministry of Education and Research, 2004), which provides guidelines for the use of ICT in the Norwegian school

system from the primary school up to the higher education levels. The Programme obliges all Norwegian educational institutions to have established a framework for the development of digital competency by the end of the year 2007. The aim here is that digital literacy will become as self-evident a skill as reading, writing and arithmetic.

### **Developing our chosen learning centre model**

UBL should, according to the University of Bergen strategic objective statement, 'contribute to the development of pedagogical resources and cooperate with academic staff in developing students' information literacy' (University of Bergen, 2005). During the process of development the UBL has been through, we have focused on the content side of the learning centre, whilst also making several physical changes in the library itself. It is how far we succeed in developing services which actually support the students' academic progress and dovetail with the efforts of the teachers responsible for individual courses that will be the ultimate measure of the efficacy of this project. Our areas of expertise are the learning centres, the digital library, and students' information literacy in collaboration with the academic staff.

The heads of the different faculty libraries did not all agree on which conclusions should be drawn from the parliamentary report (Ministry of Education and Research, 2000), and most therefore decided to postpone the implementation of changes until the University itself had instigated the reforms. However, the Psychology Library was given the task of developing a pilot learning centre, since this branch of the UBL had already recognised the potential for important new UBL services implicit in the Quality Reform guidelines, and had already begun to implement organisational changes.

### **Psychology Library – the pilot**

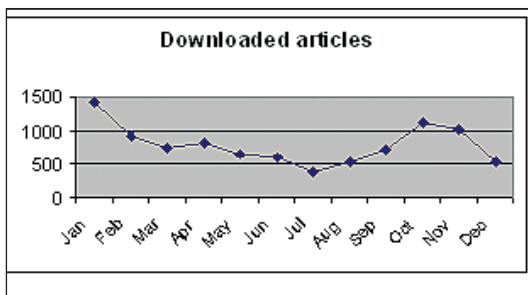
In the year 2000, the library, in cooperation with the Psychology Faculty, began developing its learning centre. Together with the UBL, the faculty granted funding for a number of new workstations for students. In order to access these, each student needed to log on to the University network using a personal password. The consequences of this move in terms of library use were very positive. There was an increase in the number of visitors, but this generated relatively little extra work for the staff, since the students generally managed fine on their own. In addition to our ordinary assistance with using reference databases etc., we provided some guidance on the subjects of logging on to and managing user

accounts, printing formats, how to relocate documents in one's allotted storage area, and so on. Students' feedback signalled satisfaction with the provision of individual workspaces, but lamented the lack of rooms for group work.

Another important factor in the development of the learning centre was the full acceptance we were granted for the principle of subscription exclusive to electronic journals. A unanimous decision of the library committee at the faculty stated that there would be no new subscriptions of any journal unable to provide electronic full-text access. Furthermore, it was decided that the library should abandon paper-based subscriptions in favour of electronic subscriptions. The committee also requested individual departments to abandon those paper subscriptions to which UBL provided electronic access.

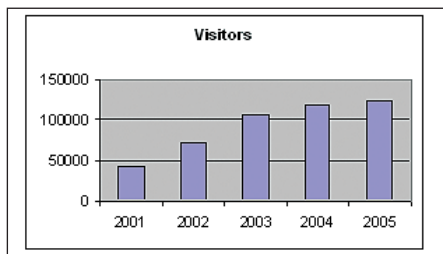
In 2002, a central provider of psychological journals offered electronic access to all its titles, with most of them providing access as far back as 1988. We were already subscribing to the paper versions of ca. 70 % of these (about 10 % of the total number of library's titles); which subscriptions were terminated. The paper issues of the volumes included in the package-subscription were stacked. In November we sent out an electronic questionnaire via e-mail to all academic staff and students at the Psychology Faculty; the number of answers were divided about equally between these groups. The most surprising and positive figure was that 75% of respondents felt they are 'often or always' able to access the electronic journals/articles they seek. About 75% gave 'not subscribed to' as the reason why they could not access the journal they wanted. Furthermore, individual comments mostly focused on suggesting titles of journals which it would be desirable to access online.

Unfortunately, there is no access to user statistics for the journal package mentioned above before April 2004. However, the 2005 figures show that twice as many articles were downloaded in the course of this year as during the final 8 months of 2004. Figure 1 shows a clear peak in downloading during the students' take-home exam period (January) and dead-line for submission of thesis (October), which may indicate that students make extensive use of electronic journals during these periods.

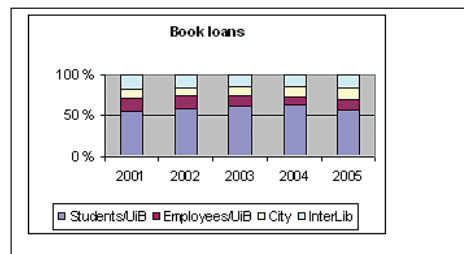


*Downloaded full text article  
from PsycArticles during 2005.  
(PsycArticles Statistics)*

In a revised budget connected with the Quality Reform, the UBL was allotted extra funding. This money was granted the Psychology Library in full to enable us to finalise the physical aspect of the learning centre. We were thus able to expand students' workspace and learning environment. At the end of 2002/ beginning of 2003 we emptied the library mezzanine of the current journal issues traditionally on display there. This area would provide group workspace, which were increasingly in demand following the shift towards group-based teaching emphasised by the Quality Reform. We also decided that all paper issues of journals in combined subscriptions (paper *and* electronic full-text) were to be stacked, and only the online full text articles would be accessible to our users, so that more shelf-space would become available in the library. Furthermore, the staff lunch-corner was replaced by a teaching corner equipped for hands-on courses. In addition, network outlets were attached to all individual readers' desks, and a WIFI network was installed. This widened access to digital library resources completed the physical basis of the learning centre.



*Visitors in the Psychology Library (UBL Statistics)*



*Student's proportional share of total book loans in the Psychology Library (Bibsys Statistics)*

The library was now in the process of becoming the kind of student-friendly higher education library required by the Quality Reform. One indicator of this was the increase in the number of visitors (Figure 2), which has kept rising steadily. Another number is the students' proportional share of total book loans (Figure 3); which shows that their percentage of total loans has remained stable at ca. 60% after we established the learning centre. These figures thus demonstrate an overall increase in library activity after our programme and the Quality Reform itself were instigated. Another figure which underlines this development is the increase in the number of information literacy courses held in the library (Figure 4).

At the beginning of the academic year 2003, UBL officially opened its first learning centre, the expressed aim of which was to develop a new kind of pedagogical resource able to support the courses given at the Psychology Faculty.

There were still a number of unfinished tasks regarding the content side of the learning centre. But at this point, the institution was ready to direct further resources towards developing a learning centre model which could increase learning and teaching quality at the University of Bergen according to the requirements of the Quality Reform.

### **Project Learning Centre 2003**

Based upon the Psychology Library's positive results as mentioned above, the UBL was allocated more funds for the further development of learning centres in the 2003 University budget. Several projects were instigated by different branches of the UBL. These funds were allocated partly towards establishing learning centres in the different faculty libraries, but also towards three common projects called Learning Centre 2003 (University of Bergen Library, 2003). One of these subprojects was the e-book access project, which is discussed by Mikki & Stangeland in this volume. The second was the Information Literacy project, which was designed to provide information literacy program through course modules; both on the internet and on campus. The background for this project was the expected increase in the number of lower-degree students needing to use the library due to the restructuring of teaching methods in the wake of the reform, involving essay-writing from the very first term onward. The third subproject was called UBL as Learning Centre; its aim was to review and recommend a learning centre model suitable for the UBL. Further in this section the aims and results of the Information Literacy subproject and the conclusion of UBL as Learning Centre review will be outlined.

Following an assessment of various established course modules and web-based instruction in information literacy, the Information Literacy project began cooperating with the University of Aalborg Library, Denmark, aiming to adapt their SWIM-programme for the UBL (University of Aalborg Library, 2002). This programme emphasises the connections between developing information literacy and academic writing, and attempts to address processes arising from this conjunction. The principal focus would not be the sources, but the students and their needs during the writing process. SWIM is based, among other sources, upon Kulthau's (2004) theorisation of the information-seeking process. For example, she demonstrates that a variety of emotions generally arise in the course of this process, and discusses how these may affect behaviour. Her research has shown a wide discrepancy between the strategies of information-seeking and processing employed by beginner and expert, respectively. The aim of the UBL Information Literacy project was to provide students pursuing lower-level degrees an information literacy e-learning programme. The website for this programme

(University of Bergen Library, 2004) was made available at the beginning of the 2004 autumn term, and was used as a basis for information literacy training in many of the courses offered by some faculty libraries. The SWIM-modules had been translated into Norwegian, and separate modules had been developed on the subjects of reference techniques and ethics as well as source criticism.

The UBL as Learning Centre review (Tønning, 2003) underlined the importance of combining form and content in the UBL learning centre model. It is vital to coordinate the resources of the traditional library, ICT, and other means of student-learning support within a pedagogical framework offering optimal overall learning conditions. The review is based on the available literature on the topic as well as insights gained through experience. Its conclusion has six main points;

- Individual and group workspaces and workstations, as well as course rooms, should be provided, and that network coverage should be made as complete as possible. These physical changes are necessary to provide an adequate learning environment for the students.
- Further development of the digital library, since this increases access to quality-controlled literature and eases the task of selecting from the multitude of all-too-easily available information.
- The guiding role of the librarian needs to be extended, and that library staff accordingly require training in information literacy and information literacy teaching.
- The necessity of systematising, planning and continuously updating cooperation with academic staff; this is described as the very foundation stone of achieving UBL's aim of introducing the learning centre as a pedagogical resource with a concrete bearing on individual courses.
- User surveys should be employed as a prominent tool in continuous quality assurances and to direct the learning centre and its supporting resources towards the real needs of users.
- Core areas and functions should be organised according to the needs of students as one primary target group, so that the competence of the UBL may be exploited across faculty boundaries.

As the coordinator of the Learning Centre project, I missed the organisational structure able to pursue and develop the services which had been made available through the project. It remained up to each faculty librarian to decide whether or not the faculty library would direct its own resources towards developing the learning centre scheme and the related information literacy courses. A decision was therefore made to develop the learning centre project further. The Learning

Centre 2003 projects, and the recommendations of the review, formed the basis of another project, Developing UBL learning centres further, called Learning Centre 2004.

### **Project Learning Centre 2004**

The aims of this project were the further development of existing information literacy courses for lower and higher degree students, to provide the SWIM-programme with instructions spoken in Norwegian, and to produce an English version of the entire course. Furthermore, the project would map the need to train staff in information literacy and information literacy teaching; this effort was in turn intended to form the basis of a general training programme for library staff.

One of the tasks of the Learning Centre 2004 project was accordingly to evaluate the autumn 2004 experiences with the information literacy programme and its relevance to actual coursework. It was concluded that revisions of the programme were needed; such as a clear framework of learning goals relevant to individual courses and degree programmes. These frameworks were integrated into a general plan to develop information literacy program for students at University of Bergen. This plan, which included suggestions as to how to integrate information literacy in students' coursework on different levels (University of Bergen Library, 2005a), was delivered to the University Education Sector Committee and the Vice-chancellor for education. The Committee discussed the plan in a December 2005 meeting, and passed it on to all faculties and placed it to local activities.

The practical impact of this plan on individual courses remains to be seen. Each faculty librarian, in association with the local sub-dean for education, here needs to actively push for the integration of information literacy with actual coursework through the faculty course committee. In this way, the concept of such integration will be more easily accepted when the academic librarian and the faculty course officer come together to plan individual programmes. In my opinion, the information literacy plan is the most important document produced by the Learning Centre 2004 project, in that it could form a basis for preserving a holistic perspective in developing information literacy in accordance with the student-activity based learning and teaching forms outlined in the Quality Reform. The plan would also be according to the emphasis in the national Program for Digital Literacy which is to be implemented within 2008.

During this project period, a UBL application was granted funds for developing the project Digital Literacy through Flexible Learning: Information searching and use of information sources in writing thesis (see Skagen & Torras in

this volume). It was agreed among these two projects and their steering committees that Digital Literacy through Flexible Learning would manage the development of the information literacy course for higher-degree students. Furthermore, the Aalborg University Library decided to produce an English version of SWIM, in the context of a general revision of the programme itself. This led to a decision by the Learning Centre 2004 project to postpone translation of SWIM into Norwegian until the new version is finished. However, the reference techniques and ethics module and the source criticism module within the information literacy course were developed further, and their translation into English is currently being carried out.

The Learning Centre 2004 project ended in October 2005. In the end report (University of Bergen Library, 2005b) the following issues were recommended;

- To use pilot groups to facilitate the integration of the information literacy programme into courses and degree programmes; and to integrate the information literacy programme and the University's learning management system.
- Internal coordination of UBL branches: the faculty libraries, the Library portal group, and the Acquisition department (which is in charge of the digital library).
- To carry out the suggested UBL staff training programme.
- UBL should formalize evaluation programme for information literacy courses offered by the faculty libraries.

So, how should one assess the implementation of the UBL learning centre model, and what lessons may be drawn from the process?

### **Challenges regarding the implementation**

A knowledge-based institution such as University of Bergen is engaged in three core areas: research, education and the public dissemination of knowledge. When major changes occur in any one of these areas, this will naturally have consequences for the UBL as well, since the library's role is to provide library and information services that will assist the University in reaching its goals.

The UBL has taken major steps towards fulfilling the requirements of the Quality Reform by means of its development of a learning centre model involving the physical adaptation of the library to accommodate more student work areas, as well as extended access to digital resources and the provision of information literacy teaching in cooperation with academic staff. However, some challenges arose during the process; these were the organisational structure, the motivation and development of skills within the librarian staff, and the cooperation with academic staff.



## **Organisational structure**

Traditionally the UBL organisational structure has been flat; nine departments at the same level directly connected to the director, in addition to an administration department. From my point of view the heads of each faculty library are responsible only for individual libraries, and are under no obligation to follow up UBL's overall strategies and yearly statement of objectives beyond their own personal interpretations of these documents. When we were to implement the learning centre projects, this became a real challenge.

The UBL is now involved in a phase of reorganisation. In my opinion, several structural operations are necessary, not least in order to increase our ability to focus on core areas and central functional aspects of the organisation as a whole; our perspective should widen beyond the confines of established faculty and departmental structures.

However, some things can be accomplished even within the current system. For example, the Director of the UBL has appointed an Education group to coordinate the implementation and further development of information literacy teaching at UBL. All visitor-serving sections are represented on this group. I think this was a wise decision given the lack of structural flexibility within the UBL demonstrated by the Learning Centre 2003 project. The Education group also serves as an example of one kind of cross-branch organisation in a core area.

The process of developing the learning centre programme and related services has been an educative one. Some experiences, such as the overwhelmingly positive feedback from students using the learning centre as their place of work, have been very good ones. However, changes also at times encounter resistance. The learning centre idea did provoke negative reactions amongst UBL staff on many levels; the level of support along the way was low in some quarters. At the Psychology Library, the first of the faculty libraries to implement major changes, this was certainly the case. The provision of new services necessitated a re-structuring of the library's priorities, and accordingly also a re-distribution of tasks. Firstly, as mentioned earlier, the emphasis on e-journals, which led to less copying work and less work regarding the journal's paper issues. Furthermore, when the learning centre – with its many computers – had been established, we received requests which required a higher level of ICT competence. Around the same time, the number of subjects and courses at the Psychology Faculty increased; this led to increased need for academic librarian work, for which purpose we were able to re-define one of our posts. Several staff members participated in the Learning Centre projects, and we employed additional temporary staff during several periods. There were considerable grey areas and differences of opinion regarding the distribution of tasks, responsibilities, and the boundaries between

different areas of competence. Re-structuring moves were made which aimed to establish a framework for how to balance and distribute tasks related to the day-to-day running of the library versus academic librarian tasks. In my opinion; in a transitional period, such changes must be strictly adhered to in order to firmly establish new routines. Thereafter, it becomes gradually more natural to work with a more flexible structure based on a mutual understanding of a clear framework of responsibility and cooperation, as is the case in the library today.

In retrospect it seems to me that the challenges faced by the Psychology Library mirror the challenges encountered in connection with the implementation of Learning Centre projects at other levels of the UBL organisational structure. When a project is given funding for, say, developing an information literacy course, this should go along with unambiguous instructions to individual branches to make use of the products developed. Because, in my opinion, the students should have approximately the same library and information services whatever subject they are studying.

In the information society of today, wherein truths are disputed as quickly as they are asserted, we need to constantly update and process new information in order to keep up. And the second challenge in term of developing library and information services is therefore to motivate the library staff continuously to develop the skills needed according to the aims in the UBL's strategic plans.

### **Motivation and the development of skills**

The UBL as an institution needs a clear strategy regarding the development of the kind of competencies that will prove necessary in the near and farther future. This in turn depends on overall long-term aims. Such a strategy will be useful in filling gaps which clearly need filling. The UBL has drawn up such a plan for 2006 based on the recommendations which emerged from the Learning Centre 2004 project, and are now in the process of carrying out the requisite staff training.

The learning centre focuses on the student and his or her learning environment. UBL intend to provide increased guidance and teaching in information literacy, both in the form of courses integrated into individual courses, and as one of the prime tasks of our visitor contact services. All staff who are in contact with students in the library should therefore be given the necessary training in information literacy and information literacy teaching. Practical pedagogical competency is required to provide courses which are in line with relevant educational theories, and it is of course vital to maintain a high level of information literacy within one's own subject area.

A major challenge during this phase will be to motivate staff to implement changes. Kaufmann & Kaufmann (2003) emphasise the importance of all parties

implied realising the nature and necessity of the changes to be undertaken so that people are able to properly coordinate their resources.

The UBL training programme emphasises the difference between the skills required of staff involved with visitor services, and those responsible for planning and conducting information literacy courses. A differentiated effort here is vital, not least because a sense of individual accomplishment is decisive in strengthening our intrinsic motivation for a particular task (Deci and Ryan, 1994, Deci et al., 2001). Staff motivation may thus be higher if a training programme is designed to enable one to adequately perform the specific tasks relating to one's actual function; as the UBL training programme is aiming to do. It is important to assimilate the pedagogical principles necessary to teach information literacy; one also needs to be able to practice such teaching skills under expert guidance. Examples of how to accomplish this are the use of work-shops and inspiration groups encouraging the best practice, guidance from experienced colleagues, and continuous evaluation of the training courses themselves (Hook et al., 2003).

### **Academic staff**

The third great challenge for the UBL in terms of developing its library and information services in the direction of increased educational relevance is clearly the systematisation of collaboration between library and academic staff. The faculty libraries have to cooperate at the administration level, the programme committee level, as well as the course and degree programme level, within their individual faculties.

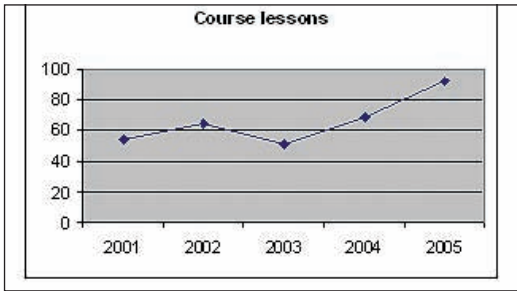
We need to actively press for the systematic integration of information literacy teaching in individual courses on different levels. As mentioned above, a UBL proposal (prepared through the Learning Centre 2004 project) for an information literacy strategy has already been discussed by the University Education Committee, and was sent on to individual faculties with a recommendation that means of local cooperation are further developed. We know by now that if the information literacy training programme for students; is to succeed, a full integration into individual courses and degree programmes is needed; this must be based on agreement between faculty library and academic staff on the subject of student needs in each course.

Many would say that academic and faculty staffs already have their hands full adapting to the demands of the Quality Reform regarding student activity based teaching, closer supervision of students, new forms of learning and assessment. It is of course reasonable to assume that some time was needed to adapt here. However, our experience in the Psychology Library with many

new subjects and courses shows that the usefulness of integrating information literacy teaching into courses is immediately recognizable. For example, one teacher stated in an e-mail after a series of courses that: 'It is quite evident from the quality of students' work who has participated in library workshops and who has not'. In several cases in all our faculty libraries, teachers have told us that our work significantly eases their job. Indeed, Bruce (1997) emphasises precisely this point: close cooperation between library and academic staff will enable the different aspects of information literacy to become fully integrated, and the teaching roles of the parties involved will be distributed in a natural way.

One local example of this is the success in integration of information literacy programme; from the lowest to the highest degree level; into the courses in health-promotion; which are singular in their use of the method of problem-based learning. The information literacy programmes on different levels are now in their curricula with its own learning objectives, and will be taught in parallel with emphasis on individual learning and text-production within the health-promotion programmes. Interestingly, this close integration process got started because library representatives happened to attend a seminar where the planning of these courses was being discussed during a break! The academic librarian was subsequently invited to participate in the entire course planning process. In this case, then, our success in integrating the information literacy program into the courses started by coincidence. However, we must go far beyond chance encounters when it comes to promoting our expertise to academic staff. Indeed, permanent features of the organisational structure facilitating and systematising such contacts should be established across the University.

We are currently promoting this kind of integration vis-à-vis other subjects areas within the Psychology Faculty, and hope that given the evidence from its pioneering success within the health-promotion course, this approach will gradually become a widely accepted part of the planning process of a range of other courses as well. The feedback from students and teachers will make very interesting reading here; this may well constitute a powerful argument for the practical value of integrating information literacy program into ever more individual courses. The information literacy courses offered by the faculty libraries are based on our course catalogue and its specific learning goals. This catalogue is in turn based on the strategic recommendations concerning information literacy developed by the Learning Centre 2004 project. Figure 4 shows a significant increase of the number of lessons provided by the Psychology library since the University implemented the Quality Reform. And the great increase in lessons held is not unique to the Psychology Library; these figures are increasing in the other faculty libraries as well.



*Numbers of course lessons given in the Psychology Library (UBL Statistics)*

The management, marketing, evaluation and promotion of library and information services, based on statistics, analyses and feedback regarding library usage is becoming an increasingly important trend in academic libraries worldwide. The UBL, too, has begun to collect such data; as they accumulate, they may be used to document the efficiency of present services and to provide arguments for the development of new ones; but the UBL needs an evaluation strategy.

When communicating the advantages of utilising the library's expertise and information literacy courses to academics, an ability to show that our services are based on ongoing empirical research may prove crucial. In my opinion, the need for such evidence based research and development at UBL should only grow in the years to come. In the case of UBL, its entire budget derives from the University; nevertheless, this budget may well be substantially increased if we are able to document how our library and information services contribute to getting students through their courses efficiently, and to providing easy access to relevant information for researchers.

## **Conclusion**

The higher education Quality Reform legitimated and allocated resources for a necessary development of UBL's services through our Learning Centre projects. The Learning Centre 2003 projects showed that the UBL organisational structure was not flexible enough to make proper use of the products developed. Today, however, at least one such structure has been established in the UBL's Education group. This will be able to function across the different branches to ensure a common profile for the different faculty libraries and to prepare the ground for cross-branch coordination of the information literacy teaching.

The UBL has increased its information literacy teaching and guidance efforts. However, even more resources will need to be directed towards these tasks in the future if we are to succeed in integrating our information literacy programmes fully into individual courses. Other factors crucial to further success

will be the establishment of channels of systematic cooperation with faculties and academic staff, and the integration of information literacy teaching into individual courses on all levels. This will require UBL staff to update their information literacy, information literacy teaching methods, guidance and communication techniques. Furthermore, services should be actively based upon evidence-based research.

### References:

- ABSON, C. (2003) The changing picture of higher education. IN OYSTON, E. (Ed.) *Centred on learning : academic case studies on learning centre development*. Aldershot, Ashgate.
- AMERICAN LIBRARY ASSOCIATION (2000) *Information Literacy Competency Standards for Higher Education*.
- BRUCE, C. (1997) *The seven faces of information literacy*, Adelaide, Auslib Press.
- COUNCIL OF AUSTRALIAN UNIVERSITY LIBRARIANS (2001) *Information Literacy Standards*. Canberra.
- DECI, E. L., KOESTNER, R. & RYAN, R. M. (2001) Extrinsic Rewards and Intrinsic Motivation in Education: Reconsidered Once Again. *Review of Educational Research*, 71, 1-27.
- DECI, E. L. & RYAN, R. M. (1994) Promoting Self-Determined Education. *Scandinavian journal of educational research*, 38, 3-14.
- FAGERLI, H. M. (2000) *En arena for læring og samhandling : om bibliotekfaglige og pedagogiske utfordringer i læringscentermodellen*, Oslo, Høgskolen i Oslo.
- HOOK, S. J., STOWELL BRACKE, M., GREENFIELD, L. & MILLS, V. A. (2003) In-house training for instruction librarians. *Research Strategies*, 19, 99-127.
- KAUFMANN, G. & KAUFMANN, A. (2003) *Psykologi i organisasjon og ledelse*, Bergen, Fagbokforl.
- KUHLTHAU, C. C. (2004) *Seeking meaning : a process approach to library and information services*, Westport, Conn., Libraries Unlimited.

- MINISTRY OF EDUCATION AND RESEARCH (2000) Frihet med ansvar : Om høgre utdanning og forskning i Norge. *Norges offentlige utredninger NOU 2000: 14*. Oslo.
- MINISTRY OF EDUCATION AND RESEARCH (2001) Gjør din plikt - Krev din rett: Kvalitetsreform av høyere utdanning [Do your duty, demand your rights: Quality Reform of higher education]. *St.meld. nr. 27 (2000-2001)* Oslo.
- MINISTRY OF EDUCATION AND RESEARCH (2004) Program for digital kompetanse 2004-2008 [Program for Digital Literacy]. Oslo.
- TONNING, A. S. V. (2003) Universitetsbiblioteket som læringscenter. Bergen, Universitetsbiblioteket i Bergen.
- UNIVERSITY OF BERGEN (2005) Strategisk plan 2005 – 2010 for Universitetet i Bergen Bergen.
- UNIVERSITY OF BERGEN LIBRARY (2003) Læringscenter 2003 [Learning Centre 2003].
- UNIVERSITY OF BERGEN LIBRARY (2004) Informasjonskompetanse - Gode informasjonssøk til oppgaveskriving [Information Literacy - Good Information Search for Writing].
- UNIVERSITY OF BERGEN LIBRARY (2005a) Strategi for informasjonskompetanse ved Universitetet i Bergen [Strategy for Information Literacy at the UiB]. Bergen.
- UNIVERSITY OF BERGEN LIBRARY (2005b) Videre utbygging av læringscenter ved Universitetsbiblioteket : Sluttrapport, oktober 2005 [Developing UBL learning centres further: End report, October 2005].
- UNIVERSITY OF AALBORG LIBRARY (2002) SWIM - Streaming Webbased Information Modules. Aalborg.





# User Education at the Digital Library: Physical and Intellectual Access to Information through Digital Literacy

by Maria-Carme Torras and Therese Skagen

## **Introduction**

The Digital Library provides physical access to a vast variety of information resources. However, giving users physical access to information does not guarantee that they become informed (Buckland 1991). If at the same time, we look upon libraries as learning centres (see Tønning this volume), the Digital Library can play an equally central role in helping them gain “intellectual access” (Buckland 1991) to information. In this chapter, we discuss how information and digital literacy education can facilitate the user’s intellectual access to information. The chapter presents how the University of Bergen Library (hereafter UBL), in collaboration with three other Scandinavian higher education libraries, has developed online learning objects for the virtual classroom to help students improve their digital and information literacy. More specifically, the learning objects have been designed to help students with their information search process and use of information sources in their writing process.

This chapter is organised as follows. Firstly we outline why and how the learning objects came about, with a focus on the pedagogical framework and learning goals underlying the production of materials. Secondly, the learning objects are described. Subsequently, an overview of the first stages of the implementation process is provided, which highlights the collaboration between library and faculty. Finally, the conclusion summarises the main points in planning digital and information literacy user education at the Digital Library.

## **Background**

A considerable number of Scandinavian academic libraries have redefined themselves as learning centres over the last years (see Tønning this volume). They do not just offer document collections, but also computers and working areas, as well as user support services and user education programmes which contribute to students’ effective learning (Fagerli 2000).

The information flow is larger today than ever. Libraries provide students

with both printed and electronic sources such as digitised documents, e-books, e-journals and open repositories. Apart from having access to a much wider range of library sources, students have become keen users of the open Internet and its search engines. The search for and use of information in our present society entail a number of challenges and needs, as identified for instance in the UNESCO Information for All Programme 2005-2006 (UNESCO 2005), The Prague Declaration “Towards an information literate society” (US National Commission on Library and Information Science and National Forum on Information Literacy 2003) and the Alexandria Proclamation on Information Literacy and Lifelong Learning (UNESCO and National Forum on Information Literacy 2005). These documents highlight the importance of digital literacy and information literacy for the information society. Information literacy allows the user to decide when and why they need information, where to find it, and how to evaluate, use and communicate it in an ethical manner (Chartered Institute of Library and Information Professionals 2005). In this way, information literacy plays an essential role in reducing the digital divide and in promoting tolerance and democratic values across cultures.

Since a great deal of information is nowadays accessed digitally, computer literacy and thus digital literacy have also become literacies of paramount importance. The Norwegian government has taken action within the education system to ensure that citizens become digitally literate. The white paper “Culture for learning” (“Kultur for læring”) (Ministry of Education and Research 2004) highlights, among the described education goals, the importance of digital literacy through the implementation of the Programme for Digital Literacy (2004-2008) at all levels of the educational system. Digital literacy is defined as consisting of both ICT skills and information literacy. Students at higher education institutions are expected to be offered education in information literacy by 2007.

The Norwegian Reform of Higher Education (Ministry of Education and Research 2001), introduced in 2003, encourages problem-based learning and a task-based approach to teaching, as well as new exam and evaluation methods. As a consequence, students are expected to write assignments and projects earlier and more often than they used to do. Students, as independent learners, need to be able to navigate in the information sea, which includes an ever increasing number of electronic resources.

The Programme for Digital Literacy and the Reform of Higher Education pose new pedagogical challenges for Norwegian academic libraries. Both library and faculty are aware of the fact that physical access to information alone does not make students information literate. Academic libraries have come to realise that they have a double goal to achieve: to make electronic resources available for students and to assist them in their learning process (Torrås & Vaagan

forthcoming). In terms of promoting digital literacy, library user education which focuses on information literacy, as defined above, will help students intellectually access the information they need and use it critically, creatively and ethically. In this way, the academic library looks upon itself as an integrating part of the higher education learning arena (Sætre 2002).

### **A User Education Programme for Information Literacy**

Since 2003, UBL has carried out a number of projects to develop a new user education programme for information literacy (see Tønning this volume). The goal of these projects has been to move towards a situated learning model where the focus is on the student's needs and the learning process is contextualised. In the new face-to-face teaching programme, students acquire information literacy as they work with their academic texts (Arnesen *et al.* 2004). This situated model calls for closer collaboration with faculty in order to be able to provide embedded library courses in the curricula. The responsibility for successful acquisition of information literacy is to be shared by the student, the academic staff and the library staff.

The situated learning model is challenging as it requires both traditional roles and new pedagogical skills from the librarian. As Kuhlthau (1994:114) argues, the librarian can adopt “[...] a variety of roles in relation to the user that may be thought of as different levels of service”. Traditionally, the librarian has acted as a *locator*, *identifier* and *advisor* (Kuhlthau 1994:116-118) in their interaction with their user. The kind of mediation that these three roles entail proves to be effective and helpful at later stages of the information searching process,<sup>1</sup> when users have a clear task focus and thus can articulate their information needs in a specific way. However, Kuhlthau notes that, at the early stages of the information searching process, the librarian can be more helpful if, as a *counsellor* (Kuhlthau 1994:118-120), they provide intervention into the process of the user. At the early stages of the information searching process, the student's cognitive state of mind is characterised by uncertainty. The librarian, like the academic supervisor, can help the student by guiding them through the creative process of constructing meaning, that is, of seeking certainty and clarity in their academic work. The librarian's counselling role is determined by the student's information needs at each given point.

UBL's face-to-face user education programme for information literacy builds upon the idea of process. The focus on learning as a process contributes to the acquisition of transferable skills and therefore to lifelong learning. Further, our user education programme encourages the librarian, as a pedagogue, to

---

<sup>1</sup> See Table 2 for a summary of the information searching process as described by Kuhlthau (2004).

consciously decide on which role to play depending on the information search stage the student is at and thus on their specific information needs at that point.

A first evaluation of the new information literacy courses by students and library staff at UBL showed that our user education programme should be better tailored to the specific needs of postgraduate students. Establishing learning goals for more advanced information literacy skills and providing a closer link between searching and thesis writing in the courses emerged as clear areas for further course development. The evaluation also showed that user education should cater for two increasing student target groups in Norwegian higher education, namely international and distance education students.

The project Digital Literacy through Flexible Learning: Information Searching and Use of Information Sources in Thesis Writing<sup>2</sup> was launched in 2005 to support students in their learning process by promoting their digital and information literacy. The project aims to develop a set of learning objects, *Søk og Skriv* (“Search and Write”), specially designed for distance education students at postgraduate level. The learning objects can be integrated in web-supported courses at higher education institutions. An English version of *Søk og Skriv* will be made available by 2007 to cater for international students.

### **The Learning Object Set: *Søk og Skriv***

*Søk og Skriv* (“Search and Write”) has been designed as a set of learning objects to promote students’ information and digital literacy as they work with their academic projects.<sup>3</sup> More specifically, students are guided through the information search and writing processes. Further, they are encouraged to reflect on the ethical, critical and creative use of information and to engage in a variety of activities that help them in the process of writing a research paper or a thesis. Searching and using information are presented as processes that go hand in hand with the writing process in the students’ wider process of constructing meaning. The searching process helps the student narrow down their research question. By the same token, a gradually more specific research question enables the student to conduct searches for more pertinent information.

---

<sup>2</sup> Digital Literacy through Flexible Learning ([www.ub.uib.no/prosj/DK/index.htm](http://www.ub.uib.no/prosj/DK/index.htm)) is a two-year collaboration project between the academic libraries in Bergen, that is, University of Bergen Library, Bergen University College Library and Norway’s Business School Library. Aalborg University Library, Denmark, also collaborates in this project. The project is funded by both Norway Opening Universities (Norgesuniversitet) and the collaborating Bergen libraries.

<sup>3</sup> *Søk og Skriv* is freely available on [www.sokogskriv.no](http://www.sokogskriv.no).

It is important that the learning object contents take into account the students' situation in their learning process. At postgraduate level, students have already acquired knowledge on academic writing and information searching through their study years. The evaluation of our new information literacy courses showed that postgraduate students had different needs from undergraduates when carrying out their academic work (see the previous section). The goal of the learning objects should thus be to help postgraduate students meet their specific needs. For this reason, one of the first project tasks was to establish a clear set of learning goals, and based on them, to design the learning object contents. More specifically, an attempt was made to distinguish learning goals for undergraduate and postgraduate students. In our process of establishing these two sets of learning goals, we were inspired by the information literacy standards developed in the United States (The Association of College and Research Libraries 2000; Middle States Commission on Higher Education 2003) as well as in Australia and New Zealand (Bundy (ed.) 2004). Accordingly, we defined information literacy as consisting of the five components described in Table 1 below, and decided on which specific components the library user education should work on in collaboration with faculty and/or the student.

| Information literacy components  | Learning Goals  |   |
|--|---|---|
|  | Undergraduate level   | Postgraduate level  |
| 1. Formulating a research question and expressing an information need              | <ul style="list-style-type: none"> <li>• To choose a topic and become acquainted with formal project guidelines</li> <li>• To formulate a research question</li> <li>• To understand that information is needed to expand one's knowledge, and to support one's ideas and opinions</li> <li>• To define a specific need for information</li> </ul>  | <ul style="list-style-type: none"> <li>• To choose a topic and formulate a research question</li> <li>• To justify topic choice through: <ul style="list-style-type: none"> <li>◦ Placing it in the context of earlier research</li> <li>◦ Considering theoretical framework and methodology</li> </ul> </li> <li>• Assessing project feasibility</li> </ul>  |
| 2. Choosing and accessing information sources. Locating and collecting information | <ul style="list-style-type: none"> <li>• To distinguish between information sources</li> <li>• To judge the appropriateness and quality of sources for the task at hand</li> <li>• To search in different information sources with appropriate search strings</li> <li>• To locate and access documents</li> </ul>  | <ul style="list-style-type: none"> <li>• To understand how information is quality assured in the publication process</li> <li>• To judge the appropriateness of the information source for the task at hand</li> <li>• To search in subject-specific sources with appropriate search strings based on special subject features</li> <li>• To evaluate search results in order to decide appropriateness and relevance to the topic</li> <li>• To change search strategy to ensure that the amount of information is sufficient to solve the task at hand</li> </ul> |
| 3. Evaluating sources critically   | <ul style="list-style-type: none"> <li>• To understand the concept of critical evaluation of sources</li> <li>• To become familiar with basic evaluation criteria</li> <li>• To recognise the relationship between good use of sources and academic quality in own one's and others' works</li> <li>• To evaluate the appropriateness of relevant sources based on own academic task</li> </ul> | <ul style="list-style-type: none"> <li>• To evaluate academic works in terms of content, context and use of sources.</li> <li>• To master the use of evaluation criteria in a variety of sources (websites, academic and popular articles, books)</li> <li>• To be familiar with and critical of peer reviewing and impact factor as quality assurance criteria.</li> </ul>   |
| 4. Using information in own academic texts   | The learning goals for this component are the faculty's and the student's responsibility.   |   |
| 5. Critical reading and academic integrity   | <ul style="list-style-type: none"> <li>• To understand what academic integrity is</li> <li>• To understand what plagiarism is and its implications</li> <li>• To do referencing in a correct way</li> </ul>   | <ul style="list-style-type: none"> <li>• To use a reference management tool for referencing and for systematising gathered literature</li> <li>• To be familiar with research ethics and the copy right law</li> </ul>  |

Table 1. Information literacy learning goals for library user education (University of Bergen Library 2005).

The learning goals in Table 1 show postgraduate students' greater needs for reflection and critical thinking in their search for and use of information. They are expected to be able to justify to a greater extent their choices of information sources and their information use in their research work. Further they are expected to use information in their academic production in a way that a contribution to new knowledge is made in their research field. Writing academic texts for knowledge dissemination in academia also raises a number of ethical issues which students should be able to tackle in a responsible way.

*Søk og Skriv* helps students deal with the information searching and research issues mentioned above. It consists of five main learning objects, which in turn are based on Kuhlthau's (2004) model of the information searching process. The learning objects presented in Table 2 cover the information searching process as well as a number of related writing actions. In addition they cover other essential information literacy components such as the creative, critical and ethical use of information. Academic writing plays a central role in students' learning process. As Dysthe (1996) puts it, learning is both an individual cognitive process and social interactive process. Language, and thus writing as a form of expression, connects the cognitive and the social learning processes.

| Learning object  | Information search process                             | Writing actions   | Other actions or strategies                     |
|--|--|---|---|
| 1. Task initiation (Oppgavestart)                                    | Task initiation (Stage 1)<br>Topic selection (Stage 2) | Brainstorming<br>Mind mapping<br>Reflection texts             | Reflecting on research ethics                   |
| 2. Get an overview (Få oversikt)                                     | Prefocus exploration (Stage 3)                         | Annotated bibliography<br>First outlines<br>Project statement |   |
| 3. Find and combine keywords (Finn og kombiner nøkkelord)            | Focus formulation (Stage 4)                            | Listing and structuring keywords                              |   |
| 4. Search, evaluate, collect and write (Søk, vurder, samle og skriv) | Information collection (Stage 5)                       | Draft writing<br>Writing for the study group                  | Critical evaluation of sources<br>Referencing   |
| 5. Closure (Avrunding)   | Search closure (Stage 6)                               | Conclusion writing<br>Final writing up                        | Ethical use of sources<br>Presenting one's work |

Table 2. Søk og Skriv : Learning object content

To illustrate the different phases of the writing and information searching processes, students are invited to follow character Oda, a distance education student, and her progress in writing a research paper. To promote the student's independent learning, each object includes activities which encourage the student to produce different text types (e.g. brainstorming and outline writing) in connection with their own thesis or research paper writing.

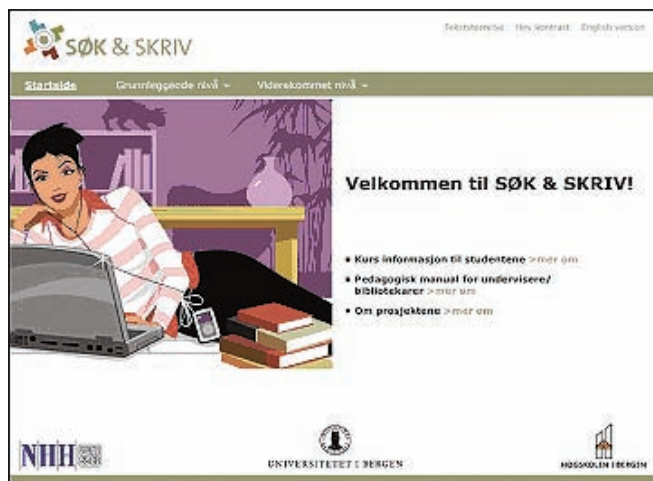


Figure 1. *Søk og Skriv* homepage ([www.sokogskriv.no](http://www.sokogskriv.no))

Furthermore, the learning objects include activities that encourage the student to reflect on the information searching process from choosing their topic and research question to formulating and combining their own keywords. The goal of these activities is to help the student become more aware of their information needs, and help them find strategies to meet those needs. The student learns by doing and reflecting (Dewey 1944), which lies at the heart of the constructivist view of learning. As Kuhlthau (2004) argues, information seeking is an intellectual process. It is important to make the student aware of how their information needs evolve from a vague awareness of an information gap and culminates in their location of information that contributes to constructing meaning. This kind of learning has transferable value, and prepares the student better for the information tasks they will encounter in their future private and professional lives.

### ***Søk og Skriv* in the Virtual Learning Environment: Collaboration between Library and Faculty**

Alongside the development of the learning objects, their integration in the university virtual learning environment needs to be worked out. At the time



of writing this chapter, two integration models have been designed. These two models will be tested through two pilot studies.

In the two pilot studies, the learning objects are integrated in the virtual learning environment of two different distance education degrees at postgraduate level: health promotion (University of Bergen) and midwifery (Bergen University College). These health promotion and midwifery degrees are actually based on a blended learning model. Students attend both virtual and on-campus classes. These two degrees highlight problem-based learning and evidence-based research. At the time the students participate in the pilot studies, they are at the initial stages of writing their thesis. The teaching which has been planned for the pilots, and where *Søk og Skriv* has been integrated virtually, aims at assisting the student in the totality of their research process by promoting both information literacy and academic writing skills.

The evaluation of the pilots will be based on a survey and a focus group interview. The evaluation of one of the pilot studies will focus on the contents and layout of *Søk og Skriv*. The other pilot will in addition become part of a larger research project to determine learning outcomes, and what practical implications they may have for the students' professional life.

Collaboration between library and academic staff is an essential condition for successful embedded digital and information literacy education. For this reason, the project Digital Literacy through Flexible Learning counts with an advisory committee assisting the project members in the tasks of developing the learning objects and planning the pilot studies as well as designing the evaluation of the learning objects. The committee consists of academic and library staff members as well as students.

The collaboration between academic and library staff members when designing the pilots has been fruitful for a number of reasons. It has been easier to find ways to achieve the established learning goals, whether they were the faculty's or the library's main responsibility. The academic staff have helped decide on what course content and timing was best suited to the students' specific learning needs. Last but not least, through this project, new communication channels have been opened between faculty and library for further collaboration. Both faculty and library share the ultimate goal of empowering the student, that is, promoting "[...] the development of independent learning skills, also known as the "learn-how-to-learn" approach, within the wider perspective of lifelong learning" (Andretta 2005:1). Our project work has provided both parties with strategies for how this goal could be achieved in a more formalised and comprehensive way. Empowering the student through digital and information literacy education has become a gradually more explicit goal. In trying to define it and discussing how to achieve it, faculty and library have become more aware

of what responsibilities each of them have in helping students learn how to learn. In this context, academic and library staff have acknowledged the value of the librarian's facilitating role of *counsellor* in the student's process of constructing meaning. In this way, faculty counts with the library as a new partner in their task to empower the student through the development of independent learning skills.

## Conclusion

This chapter has presented UBL's educational development strategies within the wider framework of building up a digital library. UBL's user education seeks to facilitate the students' navigation in the digital information sea, with the ultimate goal of empowering them through information skills that promote their lifelong learning in our current knowledge-based society. Through the project Digital Literacy through Flexible Learning, the learning object set *Søk og Skriv* has been created to improve students' digital and information literacy skills as well as academic writing skills. *Søk og Skriv* has been specially designed for postgraduate distance education students. Different models of integration of the learning objects in the students' virtual learning environment have been developed.

Collaboration between faculty and library has been a key factor for embedded teaching in a blended environment which is adequately tailored to the students' specific learning needs. The faculty and library's joint work in this project has laid the ground for further wider institutional collaboration in the task of embedding information literacy education in the curriculum. Alongside institutional collaboration, UBL needs to work further on librarian training programmes that promote their different mediating roles in their interaction with the user, and amongst them, the more challenging role of *counsellor*. It is through the pedagogical training of librarians that the digital library can be built up as an integrating part of the higher education learning environment.

## References

- Andretta, S. (2005). Information Literacy: Empowering the Learner "against all Odds" [online]. Paper presented at the LILAC Conference. 4 April 2005, Imperial College, London. Available from: <http://www.cilip.org.uk/NR/rdonlyres/FD11F35B-69DF-41D9-B0E1-38DC47BA6FEB/0/andretta.pdf> [accessed 20 February 2006].

- Arnesen, E., Lein, R., Mikki, S., Skagen, T., Tonning, A.S.V. and Torras, M.C. (2004). New Pedagogic Challenges for the University Library [online]. In *Scandinavian Public Library Quarterly*, 33, 7, 14-17. Available from: [http://www.splq.info/issues/vol37\\_3/05.htm](http://www.splq.info/issues/vol37_3/05.htm) [accessed 3 February 2006].
- Association of College and Research Libraries (2000). *The Information Literacy Competency Standards for Higher Education* [online]. Chicago, Illinois: American Library Association. Available from: <http://www.ala.org/ala/acrl/acrlstandards/standards.pdf> [accessed 3 February 2006].
- Buckland, M. (1991). *Information and Information Systems*. New York, Greenwood Press.
- Bundy, A. (ed.) 2004. *Australian and New Zealand Information Literacy Network. Principles, Standards and Practice* [online]. Available from: <http://www.caul.edu.au/info-literacy/InfoLiteracyFramework.pdf> [accessed 6 February 2006].
- CILIP Information Literacy Group (2005). Defining Information Literacy for the UK [online]. *Update Magazine*, Jan/Feb. Available from: <http://www.cilip.org.uk/publications/updatesmagazine/archive/archive2005/janfeb/armstrong.htm> [accessed 28 April 2005].
- Dewey, J. (1944). *Democracy and Education*. New York, Macmillan.
- Dysthe, O. (1996). "Læring som dialog" – Kva inneber det i høgare utdanning? In Dysthe, O. (ed.) *Ulike perspektiver på læring og læringsforskning*. Oslo, Cappelen Akademisk forlag, p.105-135.
- Fagerli, H. M. (2000). *En arena for læring og samhandling: om bibliotekfaglige og pedagogiske utfordringer i læringscentermodellen*. HiO-rapport, 4. Oslo, Høgskolen i Oslo.
- Kuhlthau, C. C. (2004). *Seeking Meaning. A Process Approach to Library and Information Services*. 2<sup>nd</sup> edition. Westport, Connecticut, Ablex.
- Middle States Commission on Higher Education (2003). *Developing Research & Communication Skills: Guidelines for Information Literacy in the Curriculum*. Philadelphia, Pa, Middle States Commission on Higher Education.
- Ministry of Education and Research (2001). *St. melding nr. 27. (2000-2001). Gjør din plikt – Krev din rett. Kvalitetsreform av høyere utdanning* [online]. Available from: <http://odin.dep.no/ufd/norsk/dok/regpubl/stmeld/014001-040004/dok-bn.html> [accessed 2 February 2006].
- Ministry of Education and Research (2004). *St.melding nr. 30. (2003-2004) Kultur for læring* [online]. <http://odin.dep.no/kd/norsk/dok/regpubl/stmeld/045001-040013/dok-bn.html> [accessed 2 February 2006].

- Sætre, T. P. (2002). *Biblioteket som læringsarena med bibliotekaren som pedagog : Et utviklingsarbeid knyttet til bibliotekarutdanning med sikte på å etablere et nytt fagområde kalt 'bibliotekpedagogikk'*. Høgskolen i Bergen Skriftserie, 3.
- Torras, M.C. and Vaagan, R.W. (forthcoming). Websites and Internationalisation: A Survey of Norwegian Academic, Research and Special Libraries. To appear in *Libri*.
- UNESCO (2005). *Information for All Programme (IFAP)* [online]. Available from: [http://portal.unesco.org/ci/en/ev.php-URL\\_ID=1627&URL\\_DO=DO\\_TOPIC&URL\\_SECTION=201.html](http://portal.unesco.org/ci/en/ev.php-URL_ID=1627&URL_DO=DO_TOPIC&URL_SECTION=201.html) [accessed 3 February 2006].
- UNESCO and National Forum on Information Literacy (2005). *The Alexandria Proclamation on Information Literacy and Lifelong Learning* [online]. Available from: <http://www.ifa.org/III/wsis/BeaconInfSoc.html> [accessed 3 February 2006].
- University of Bergen Library (2005). *Undervisning i informasjonskompetanse: Fastsettelse av generelle læringsmål og innhold for UB-kurs etter studienivå* [online]. Available from: <http://www.ub.uib.no/prosj/LS2004/alleniva-laeringsmal15.10.05.doc> [accessed 20 March 2006].
- US National Commission on Library and Information Science and National Forum on Information Literacy (2003). *The Prague Declaration. "Towards an Information Literate Society"* [online]. Available from: <http://www.nclis.gov/libinter/infolitconf&meet/post-infolitconf&meet/PragueDeclaration.pdf> [accessed 15 December 2005].

# *Bibliotekportalen - the Library in Cyberspace*

By Pål H. Bakka

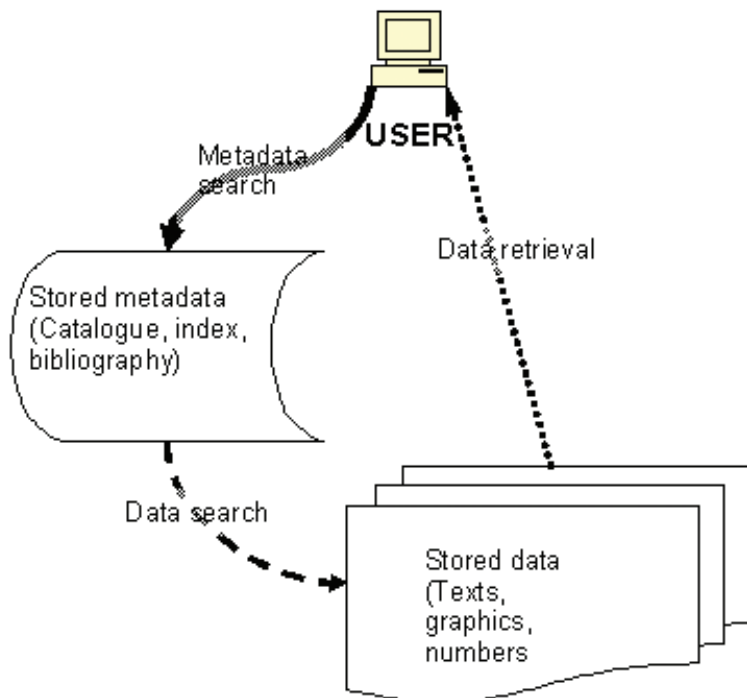
## **Why a Portal?**

There are two ways of looking for information. Searching and browsing. Searching is looking for something you know is there. Browsing is “unstructured searching” – looking for something somewhere that may or may not be there. In the library context a “search” means looking up metadata in a catalogue, index or bibliography. Browsing is going directly to the shelves. The joy of searching is finding what you’re looking for. The joy of browsing is finding something you really need.

Browsing is, needless to say, frowned upon by real librarians. Open stack libraries and good subject indexes facilitate browsing; which may be why they are relatively new to the Norwegian academic library system.

Schematically searching looks like this:

Fig. 1: Librarians search:  
The search process

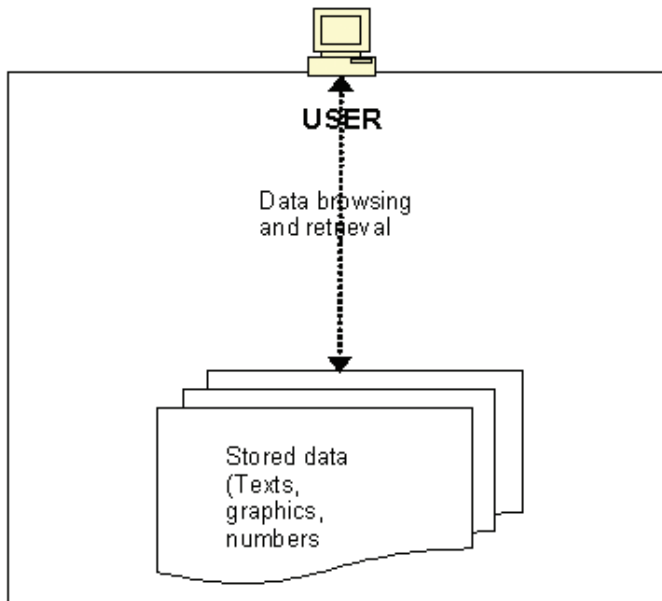


The traditional library search is, using modern information science terms, a “metasearch” – searching for data about data. Conventionally the kind of data found in catalogue entries describing the objects containing the information (data) needed. Metadata searching never identifies the entire amount of information contained in the described documents. The librarian’s rule of thumb is that half the information stored in a collection will never be found using the catalogue only ; which is why users prefer browsing for information in the data sources themselves by going directly to the stacks or the reference works.

Libraries have always offered facilities for browsing by organizing its data holding objects by content, book collections arranged by subject, or systematic catalogues or subject indexes.

Schematically browsing looks like this:

Fig. 2: User browse: The direct data search



The difference between searching and browsing is that the latter is performed directly on the information itself without going by way of predefined metadata; increasing the chance of finding useful information even where a metasearch has indicated a very low chance of success. It may lead the user into very fruitful digressions or make him waste his time in searching for something not there. The major problem with browsing is that users actually are satisfied with what they get and stop looking. Finding what they think they need, they are satisfied.

Even if what they find is not the best available information. Browsing the shelves is, after all, infinitely more satisfying than searching the catalogue. Both are necessary in the process of procuring scientific information. But browsing is in fact vital in establishing the knowledge necessary to perform informed metadata searches. Uninformed searching will always be afflicted by GIGO<sup>1</sup> and may be a total waste of time.

### **Why a Portal**

Twenty-odd years of academic librarianship at the University of Bergen Library – most of them spent in the Social Science Faculty Library - has lead the author to the conclusion that Norwegian academic librarianship really is about logistics. A library is a warehouse, its task document delivery. The core library activity is the production of metadata used to retrieve the documents stored. The only unique “data” produced by a library in its cataloguing efforts, the minute biographical document description, is the shelf signature that indicates exactly where a document is found on the shelves. The sheer bulk of the catalogue rules indicate that, in libraries, cataloguing rules.

The information stored in the warehouse has never been subject to the same attention. The effort put into subject indexing or opening the collections to browsing has, at least in the University of Bergen Library, been far less. Ensuring data quality as opposed to metadata quality – or data access as opposed to document access – never was a major concern. Not because it was considered unimportant. It was not regarded as a problem. Quality was ensured through collection development, the users of the small elite university knew their needs. The library was inwards looking to the where my old professor, Stein Rokkan, lambasted the University of Bergen Library as a “book mausoleum”.

Since “librarians search” in order to deliver documents, metadata are quite sufficient for the library as an institution. That users don’t find is a minor problem compared to the chaos ensuing if metadata quality suffers. Library logistics are immensely complicated because libraries are extremely complex warehouses containing literally millions of unique objects. Like any good storekeeper a librarian should properly be more concerned with keeping order in the warehouse than with the use to which the objects are put. Dissemination of knowledge, research or education for information literacy is at best auxiliary activities compared to the maintenance of collections and document delivery systems.

---

<sup>1</sup>“Garbage in, garbage out.” It refers to the fact that computers, unlike humans, will unquestioningly process the most nonsensical of input data and produce nonsensical output”. ([http://en.wikipedia.org/wiki/Garbage\\_in,\\_garbage\\_out](http://en.wikipedia.org/wiki/Garbage_in,_garbage_out), Mar. 14th 2006)

That being said, the major priority of any library has always been prompt delivery of documents, for preference from its own collection. And as the good shopkeepers we are, librarians have always had a keen interest in the contents of the warehouse. We like to know what is hidden on our shelves. We need to know in order to provide our customers with documents containing the information they need. And we know better than anybody else that it is there – somewhere.

The existence of the World Wide Web means that “somewhere” no longer need be a shelf in some library. It’s out there on the Web. Which means it’s in there – three clicks away.

Or should be. In an ideal world.

The Internet contains everything. WWW really means the Whole Wide World. More prosaically, it contains a lot of unstructured and ill described data in addition to the minute quantity of metadata provided by libraries. At the same time these accumulated metadata are the only quality-controlled bits of information freely available on the Web. The rest has not and never will be subject to quality control. Searching the Web is best left to search engines. But it is a browser’s paradise.

To an academic library the Web is an opportunity. To its parent institution it is equally a threat. The Web is information overload come true. You can browse for years and not find a single piece of relevant information. If students do not drown in the ocean of raw data, they may come up with erroneous data or worse. They may even use it for cribbing – just like they used to use books.

For libraries the challenge of the Web is dual:

- To integrate their metadata with the data available on the Web in a system that allows simultaneous searching and browsing.
- To ensure the quality of the data made available to their users.

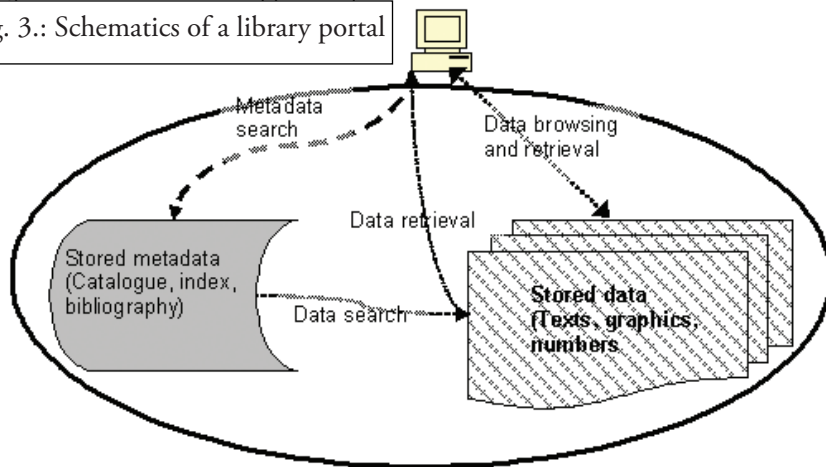
To meet the first challenge libraries world-wide have adopted gateway or portal technologies.

A portal is an interface that integrates metadata searching with Web browsing. It is a search engine in which the components of the virtual library are presented, indexed and made cross-searchable as an integrated whole; usually linked to an automated document delivery system that enables all users to download documents directly.

Schematically a library portal looks like this:



Fig. 3.: Schematics of a library portal



The diagram looks very neat and tidy – as is a functioning portal. Getting a portal to function is not as neat and tidy. Libraries should, as always, heed Napoleon’s dictum “*tous est dans l’execution*” – everything is the execution.

### Implementing a Library Portal: Institutional Constraints

It looks very easy on the Web. Enter any word into the search window and any search engine will come up with thousands of results. But only hits from Web pages indexed – made searchable in the search engine used – will be found. The major problem confronted when implementing a library portal is that the content libraries need to present comprise an aggregate of digitized information sources dating from various stages in the development of digital information systems.

There are four major classes of information sources that must be cross-searchable in a library portal. By historical order of appearance these are:

- Metadata collections: Library catalogues, subject and article indexes, bibliographies
- Data archives storing statistical information in numerical form
- Digitized text databases, spanning from collections of legal, technical and official documentation through electronic books and journals to reference works
- Web-pages

The coming of digital electronic information technology in library world was very gradual and very smooth. The first steps were taken by commercial metadata

publishers. Excerpta Medica launched its first database, Embase, in 1960<sup>2</sup>. Even more important was the work done by bibliometric visionary Eugene Garfield who in 1960 founded the Institute of Scientific Information that began publishing the Science Citation Index in 1964<sup>3</sup>. The Social Sciences Citation Index went online in the Dialog system in 1972<sup>4</sup>. In Norway, however, the introduction of digitized external databases accessed through dial-up connections did not change the way libraries searched. The introduction of the Bibsys online catalogue in the early 1980s neither changed cataloguing nor searching nor document delivery. It just made everything far more efficient. More documents could be catalogued, found and delivered. The introduction of the Z39.50<sup>5</sup> protocol in the late 1980s enabled online cross-searching of library catalogues, revolutionizing interlibrary lending operations. Libraries still processed metadata and the systems merely made things easier.

The development of digitized data archives such as the NSD on the other hand was undertaken by professional associations and statistical institutions without the library system being involved at all<sup>6</sup>. Digitized texts have been around ever since Michael Hart embarked on Project Gutenberg in 1971<sup>7</sup>. Newspaper text archives emerged in the 1980s as a by-product of the introduction of a new technology in that industry<sup>8</sup>. Specialized text databases serving special professional needs, such as *Lovdata*<sup>9</sup> emerged at the same time. Large scale, and more importantly, commercial, production of digitized texts only began with the advent of cheap electronic word-processing technologies in the early 1990s.

The Major problem confronting any library is Space. While active collections development will husband Space, the time will inevitably come when the information warehouse is full and no further acquisitions can be made without seriously compromising the integrity of the collection. This is particularly so with periodicals – the putative lifeblood of scientific communication. While librarians may wonder whether this actually is the case, “ours is not to wonder

---

<sup>2</sup> ([www.reedelsevier.com/index.cfm?articleid=113](http://www.reedelsevier.com/index.cfm?articleid=113)). Now owned by the publishing conglomerate Reed-Elsevier - the most hated corporation in the library world.

<sup>3</sup> ([www.garfield.library.upenn.edu/overvu.html](http://www.garfield.library.upenn.edu/overvu.html))

<sup>4</sup> ([http://en.wikipedia.org/wiki/Citation\\_index](http://en.wikipedia.org/wiki/Citation_index)), ISI and Dialog were bought by information science giant Thomson – the second most hated corporation in the library world - in 1992

<sup>5</sup> (<http://en.wikipedia.org/wiki/Z39.50>)

<sup>6</sup> NSD – The Norwegian Social Science Data Archive – stores survey and ecological data for research purposes. It was established in 1971 and is part of the Inter-University Consortium for Political and Social Research. ([www.nsd.uib.no/english/PresetasjonE%27032.pdf](http://www.nsd.uib.no/english/PresetasjonE%27032.pdf)). The paradox is that such archives store the digitized versions of data libraries hold in paper form.

<sup>7</sup> [www.gutenberg.org/about/history](http://www.gutenberg.org/about/history)

<sup>8</sup> The Norwegian Atekst-service owned by the Schibsted Publishing House, now run by Retriever Norge AS, thus has archives going back to 1984. ([/www.retriever-info.com/about\\_us.php?section=about&subsection=1&lang=en](http://www.retriever-info.com/about_us.php?section=about&subsection=1&lang=en))

<sup>9</sup> ([www.lovdato.no/info/lawdata.html](http://www.lovdato.no/info/lawdata.html))

why". We do not know whether they really earn their keep to justify the expenditure of 90 % of our acquisition budgets. Ours is just to deliver the documents. But given the proclivity of journal publishers to respond to the imperatives of publish or perish by launching more journals and increasing the number of pages, by the 1990s academic libraries were literally meeting the wall; running out of shelf-space. No wonder the first major digitized journals archive, the JSTOR, was library initiated<sup>10</sup>.

The online journal was a godsend that saved the academic library from itself<sup>11</sup>. But it also meant that for the first time in its history a sizeable and growing part of its collection was as accessible in its original form, as data, as it previously had been as metadata, catalogue entries.

The technology needed to disseminate the online journal, the World Wide Web, was only invented in 1989<sup>12</sup> and launched as a commercially available technology in 1994-95; coinciding with the launch of the first major e-journal collections by commercial publishers such as Elsevier, Springer and Bell & Howell<sup>13</sup>

The WWW is any reference librarian's dreams come true. It provides immediate and unlimited access to information, access to far more such sources of information than any academic library could dream to hold in paper form and it makes reference materials as easily accessible to the patron as it used to be to the experienced librarian. More prosaically the Web enabled libraries to organize single interfaces for their disparate databases and replace print and CD-ROM resources with online as well as to save space by replacing official publications and statistics in printed form with links on their websites. The explosive growth of the Web as the chosen means for the dissemination of information also meant that the commercial supply of online reference sources skyrocketed.

But the supply of free sources of information increased even more dramatically. This was, in part, a good thing. Official information, publications from international organizations and from research institutes as well as gray material that had taken days and weeks of painstaking work to verify and localize was there for the downloading<sup>14</sup>. But so was everything else. By 2000 the spectre of information overload seen on the horizon in the 1980s had become real. Ninety percent of everything is crap. Reflecting on the content of the Internet

<sup>10</sup> ([www.jstor.org/about/background.html](http://www.jstor.org/about/background.html))

<sup>11</sup> There have been worries about archive access. In the case of an irreparable breakdown global telecommunications, access to back volumes of Social Text will be the least of this author's worries.

<sup>12</sup> ([www.w3.org/Consortium/history](http://www.w3.org/Consortium/history))

<sup>13</sup> ScienceDirect ([www.sciencedirect.com/](http://www.sciencedirect.com/)), SpringerLink ([www.springer.com/sgw/cda/frontpage/0,11855,5-117-2-140681-0,00.html](http://www.springer.com/sgw/cda/frontpage/0,11855,5-117-2-140681-0,00.html)) and ProQuest ([www.proquestcompany.com/about/history.shtml](http://www.proquestcompany.com/about/history.shtml))

<sup>14</sup> It is not a tale of universal progress, however. The Norwegian parliamentary records online are less easy to search and use than the paper version.

one is tempted to revise that number upwards. But the world is wide, and the Web is but a window on the world.

For an academic library this poses a problem. While we are in the business of storing information we also sift information in our decisions on collection development. While we will produce any document on demand, we will not store all of them. Standards of academic integrity apply. We will after all, never knowingly recommend a bad book and we do have a responsibility to ensure that the information we deliver is factually correct. With the Web the library responsibility for ensuring quality of content inherent in any collection development decision suddenly expanded to include the Whole Wide World. At a more practical level it should be remembered that while librarians search, patrons find. Commercial search engines very often lead patrons astray by producing a virtual fog of information in generating a five digit numbers of hits. And most commercial search engines do not index the resources libraries pay for.

Each of the components required in a library portal has its origins in distinct and different institutional requirements. The online catalogue is the product of the automation of library operations. Its strong point is that it is a structured catalogue that in most cases can be searched using a single protocol. The online journal is a response both to the real need of academic libraries to save space and increase journals usage, the need of the academic community at large for better means of scientific communication and of the publishing business to make money. Online and virtual reference sources, on the contrary, represent incremental adaptation of preexisting material forms to new material types; i.e. print or CD-ROM to online. The content of the Web is a byproduct of the incremental technological adaptation of institutions and persons using the Internet to publish the products of their own activities. The Web was a technological leap; a fanzine, however, remains a fanzine.

Due to their disparate institutional origins and age, each of these types of digitized materials present special problems for the implementation of a library portal. In particular the “legacy” databases, those well established prior to the invention of the WWW, present problems not easily overcome. The grandfather of them all, the Web of Knowledge, was not meant to be cross-searchable when launched as the Science Citation Index. The same applies to all commercial databases launched prior to 1998. The only legacy systems that are cross-searchable are the library catalogues. That is because libraries always have pooled and exchanged metadata in union catalogues because all libraries require other libraries in order to deliver the documents their users need.

The rule of thumb is that the further removed a source of information is from the library world, the more intractable it is to configure in a library portal. And in order to be cross-searchable, resources must be searchable in the portal

interface. If not, they can only be accessed in their own, “native”, interface on the web – which of course defies the object of the entire exercise: To enable searching across a range of disparate information sources using a single interface. If not a library portal is merely a glorified link collection.

It should come as no surprise that library produced metadata collections are the easiest to configure in a library portal. Neither should it be surprising that electronic text collections from major commercial publishers are among the resources tailored for portals. The most intractable are the “legacy” databases<sup>15</sup>. These are a major problem to the libraries because they very often are the mainstays of scientific communication in their fields. But a lot of commercial information sources are not easily integrated into a portal. Web pages very often can only be configured using screen-scraping techniques.

But that is just what we ought to have expected. The WWW is not a library.

It can be argued that the problems libraries confront in attempting to integrate Web resources in their portals is of their own making. Library standards for document description and control are extremely high. At the very least adapting to more widely used commercial standards, even if not of the accustomed quality, is the way forward towards better integration of digitized information provided by institutions more than one step removed from the library world.

### **Fragments of the History of the Digital Library in Norway**

The development of the Norwegian automated academic library system, Bibsys<sup>16</sup>, started in the Trondheim university library system in 1972. The University of Bergen Library joined in 1980, becoming the fourth participant in the system. After overcoming the inevitable teething troubles of going digital, the University of Bergen Library by 1995 was a fully automated and integrated university library with a reputation for efficient document delivery and very high cataloguing standards. It was expanding rapidly from a centralized closed stack library into a decentralized open-stack one; happily looking forward to getting its collections even closer to its users.

Then somebody invented the Internet.

The first inkling of things to come came with the introduction of a Windows interface with Internet connection for Bibsys III in 1994<sup>17</sup>. Searching

---

<sup>15</sup> In the Metalib Central Knowledge Base the Web of Science has its own unique configuration that must be manually updated whenever the publishers make a change in the interface. Others defy all attempts at configuration and cannot be cross-searched.

<sup>16</sup> Bibsys started as a library automation project in the Trondheim University Library System in 1971. ([www.bibsys.no/om/bibsys-status-e.htm](http://www.bibsys.no/om/bibsys-status-e.htm)) The University of Bergen Library became a member in 1980.

<sup>17</sup> BIBSYS-III - hva er nå det? and Nye søkemuligheter In Bibsys Nytt 1994-1 ([www.bibsys.no/](http://www.bibsys.no/))

Bibsys on the Web was a reality in 1995<sup>18</sup>. In 1996 Bibsys became the host for the ISI-databases<sup>19</sup>, enabling article searches using the regular library interface. The Z-search functionality bundled in Bibsys enabled searching in foreign OPACS as well.

The Web and the e-journal make their first official appearance at the University of Bergen Library in 1998, with the board minutes making the first references to the Intranet<sup>20</sup> and the first consortium for electronic journals<sup>21</sup> and formal guidelines for uploading library Web pages<sup>22</sup>. In 1999 Web resources started appearing in some numbers, with 700 electronic journals available<sup>23</sup>. The need for a portal solution was readily apparent from the very beginning. The Web is a disorderly place.

Schemes abounded. The Bibsys subject gateway made its debut in 2000<sup>24</sup>. In 2001 the National Electronic Library was the buzzword<sup>25</sup>; in 2003 the Norwegian Digital Library<sup>26</sup>.

The Norwegian library system is a two-headed monster. The academic libraries are part of the system of higher education that answers to the Ministry of Education. The public libraries with the National Library at the apex answer to the Ministry of Culture. "Library policy" is thus under the Ministry of Culture, even when it affects the academic library sector. The Ministry of Education is sublimely indifferent to its libraries; regarding them as the responsibility of their parent institutions. The absence of any central direction means that the different segments of the Norwegian library sector tend to act alone. Bibsys proved incapable of meeting the requirements of the big university libraries in accommodating their wholesale transition from print to electronic resources, just like the library bodies under the Ministry of Culture failed to meet the challenge of web-publishing. The visions of a national digital library receded in the face of the university libraries' commitment to their parent institutions - and owners. This became all the more pressing as the reform of Norwegian higher education

---

[bibnytt/94-1/bnytt941.html#biii](#)).

<sup>18</sup> World Wide Web and Post frå fansen In Bibsys Nytt 1995-1 ([www.bibsys.no/bibnytt/95-1/bnytt951.html#worl](http://www.bibsys.no/bibnytt/95-1/bnytt951.html#worl))

<sup>19</sup> ISI-databasen in Bibsys Nytt 1997-1-Mars ([www.bibsys.no/bibnytt/97-1/97-1.htm](http://www.bibsys.no/bibnytt/97-1/97-1.htm))

<sup>20</sup> Minutes of Staff Meeting, Feb.23 1998 published on the Web. University of Bergen Library Board Minutes, S20/1998a ([www.ub.uib.no/felles/organ-styr/styre/1998/sr980402.htm](http://www.ub.uib.no/felles/organ-styr/styre/1998/sr980402.htm))

<sup>21</sup> The offer in question was for Swetsnet. University of Bergen Library Board Minutes, S50/98d ([www.ub.uib.no/felles/organ-styr/styre/1998/sr981209.htm](http://www.ub.uib.no/felles/organ-styr/styre/1998/sr981209.htm))

<sup>22</sup> Proposed guidelines for UB's Web pages. University of Bergen Library Board Minutes, S10/1999.

<sup>23</sup> University of Bergen Library. Annual report 1999. Sec 3.9. Vidareutvikling/effektivisering av bibliotektenestene, med særleg vekt på elektroniske tenester ([www.ub.uib.no/felles/dok/aarsmelding1999-UB.html#\\_Toc487611074](http://www.ub.uib.no/felles/dok/aarsmelding1999-UB.html#_Toc487611074))

<sup>24</sup> Bibsys Emneportal in Bibsys Nytt 2000-1-Mars ([www.bibsys.no/bibnytt/00-1/](http://www.bibsys.no/bibnytt/00-1/))

<sup>25</sup> [www.ub.uib.no/prosj/nuu/dokument/ELF\\_sluttrappport.pdf](http://www.ub.uib.no/prosj/nuu/dokument/ELF_sluttrappport.pdf)

<sup>26</sup> University of Bergen Library Board Minutes, S02/2003f. The link to the report published by *ABM-utvikling* is rotten.

in 2001 for the first time required the university libraries to actually concern themselves with teaching activities<sup>27</sup>.

While the university libraries in 2003 reached agreement in principle with Bibsys and the national professional body, *ABM-utvikling*, on developing a portal<sup>28</sup>, they were also open for other possibilities. This led to the consortium of the four “old” Norwegian university libraries inviting tenders for a library portal from other suppliers, and in June 2004 the major Norwegian university libraries decided to accept the offer from Bibits for the Metalib/SFX-system produced by Ex Libris<sup>29</sup>.

### Implementing *Bibliotekportalen*

So far as it goes the Metalib portal represents “tried and tested” technology. It is in use at several American and British institutions and has been adopted for the Finnish national library portal *Nelli*. It consists of a search and registration program/engine, Metalib, with a Central Knowledge Base of preconfigured (ready indexed) resources and programs for configuring Web-resources (screen-scraping), and a program for registering, searching and retrieving data from electronic journals, SFX<sup>30</sup>.

Four Norwegian university libraries, Oslo, Bergen, Trondheim and Tromsø, are currently members of the portal “consortium”. As each library has its own interface and its own resources, the joint activities of the consortium are limited to sharing the same technical host, the USIT at the University of Oslo, and cooperating very closely in solving the inevitable problems encountered in implementing and running what remains brand new technology. No data system ever ran smoothly on the first try. There are still bugs to be ironed out and updates to integrate. As each update and patch creates new problems the Portal will never, ever, be “finished”.

To implement the portal at the University of Bergen Library a working group under the very capable leadership of Senior Academic Librarian Dr. Hege Folkestad was set up in the autumn of 2004. The group consisted of seven persons drawn from all of the departmental libraries representing all its professions<sup>31</sup>. The

<sup>27</sup> Cf. Action plan for ICT-supported learning in the new study programs. [www.ub.uib.no/felles/organ-styr/styre/2003/S09-2003-h.doc](http://www.ub.uib.no/felles/organ-styr/styre/2003/S09-2003-h.doc) University Library of Bergen Board Minutes, S09/2003h [www.ub.uib.no/felles/organ-styr/styre/2003/sr030403.htm](http://www.ub.uib.no/felles/organ-styr/styre/2003/sr030403.htm)

<sup>28</sup> University Library of Bergen Board Minutes, S15/2003f,i,j.

<sup>29</sup> University Library of Bergen Board Minutes, S21/2004e.

<sup>30</sup> For information on the Ex Libris portal solution, cf. [www.exlibrisgroup.com/](http://www.exlibrisgroup.com/)

<sup>31</sup> The members of the implementation group were Dr Hege Folkestad (Science Library), Ms. Regina Kufner Lein (Medical Library), Ms. Britt-Inger Bjørsvik (Law Library), Ms. Hilde Wedvich (Psychology Library), Ms Anne B. Aasmul (Arts & Humanities Library), Mr Stephen Olson (Acquisitions department), Mr. Leif Magne Iversland (IT-department) and Dr. Pål H. Bakka (Social Science Library).

author is proud to have been a member of this select group.

The task assigned to the group was very simple: To create a library portal for the University of Bergen. This entailed naming the beast, the local interface, deciding on and configuring the content in the portal, including registering several thousand electronic journals in SFX, and not least, since the University of Bergen Library chose *nynorsk* as the default language of its portal, doing the pioneering work of translating the entire portal into that language; the latter a task never undertaken at any Norwegian library. The target date for the opening of the portal was set to May 3<sup>rd</sup> 2005.

After some discussion in the implementation group and an open contest the University of Bergen Library Portal was officially named *Bibliotekportalen* (The library portal).

In retrospect the implementation went smoothly. Configuring the international resources present in the CKB – and thus the majority of the important information sources – went as smoothly as could be expected. In retrospect the library ought to have been prepared for the problems encountered in trying to configure legacy databases and national Norwegian information sources. As the Norwegian legal information system is based on such databases this was, and remains, a serious problem that awaits solution. But it meant that the intended and hoped for full integration of online legal sources in the portal was not achieved. Given the very high quality of the preexisting web-based legal information system embodied in *Juridisk nettviser* this is less of a problem than a beauty spot, however. But similar problems were encountered with other and newer, Norwegian sources such as the key electronic periodicals database Idunn.no<sup>32</sup> and the newspaper archive produced by Retriever Norway AS<sup>33</sup>. Sources with a direct institutional link to the global library system and with an international audience were easy to configure. The others were more of a challenge. It should be kept in mind that configuring a Web-resource for searching in Metalib is sometimes as time-consuming as cataloguing an intractable old book. A decision was made early to configure the Bibsys Subject gateway content in the portal. This means that several thousand web pages hide behind the 500-odd single resources in *Bibliotekportalen* at the time of writing

The configuring the electronic journals benefited from the fact the large suppliers like ScienceDirect, JSTOR, Blackwell, Proquest and Springer, are international and that libraries are their major customers. But beyond the English language

---

Valuable work was also done by the two project assistants Ms. Kristine Fjone Godal and Mr. Jan Haugseth. In the summer of 2005 Ms. Aasmul was replaced by Ms Ellen Solvik as the member for Arts & Humanities Library.

<sup>32</sup> The Idunn database is the technically worst e-journal interface on the global market.

<sup>33</sup> Both Retriever Norway AS and Universitetsforlaget, publisher of Idunn.no, have indicated a willingness to meet library-compatible standards in their upgrades.



oligopoly the e-journal business is fragmented. The configuring of say, French e-journals, remains a challenge. The portal therefore reinforces the dominant position of English as the international academic language. The problem of data exchange between SFX and Bibsys has not yet been solved either. At the time of writing 14572 journals are available *Bibliotekportalen*.

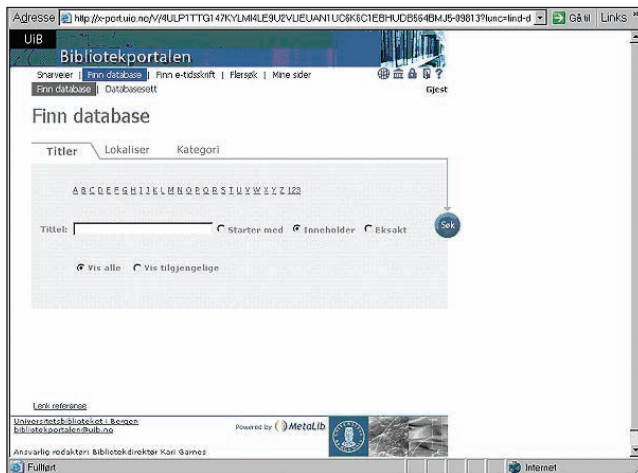
The strength of the *Bibliotekportal* lies in the system's ability to retrieve articles from electronic journals using the SFX link server. The ability to access any article of any subscribed journal regardless of where on the Web it is found justifies the entire investment. The potential for cross-searching databases remains hampered by the fact that too many of them are impossible to configure because not intended to be searched in any other engine than, say, Google – if at that.

## The Portal Described

The author would really have liked an animated icon – a heavy double door opening – for the *Bibliotekportal*. Though the icon that was designed by the implementation group is excellent:



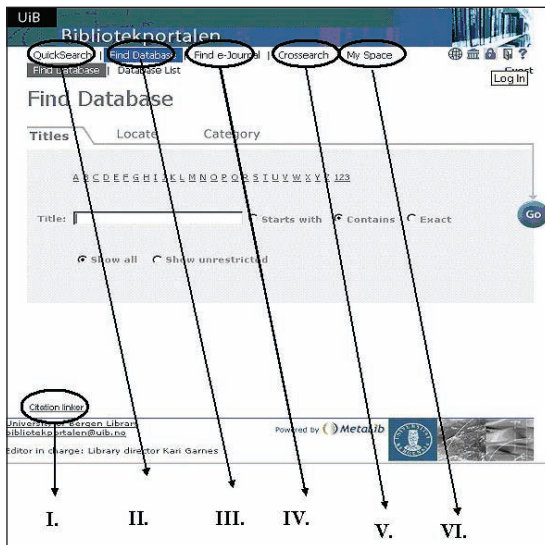
Clicking the icon brings up this screen:



The visitor is always greeted as “guest” – as indicated by the name in the upper right corner. Any person surfing the Web can use the portal to find the information the library has organized in the portal. The guest, however, will not be able to read and download resources that the host, the University Library of Bergen,

pays for. To do so the guest must log in clicking the padlock icon. The user exits the portal by clicking the door icon. Clicking the globe icon allows her to change the interface language. The default value is *nynorsk*. Clicking the temple icon allows a user to enter the portals of the other consortium members, even if Web access to licensed resources is based on the IP-address of the surfer's PC, the log-in facility enables the use of personalized features in the portal. It is thus possible for the library to tailor the content of the interface, the predefined QuickSearch sets and contents of My Space, for the need of any user group.

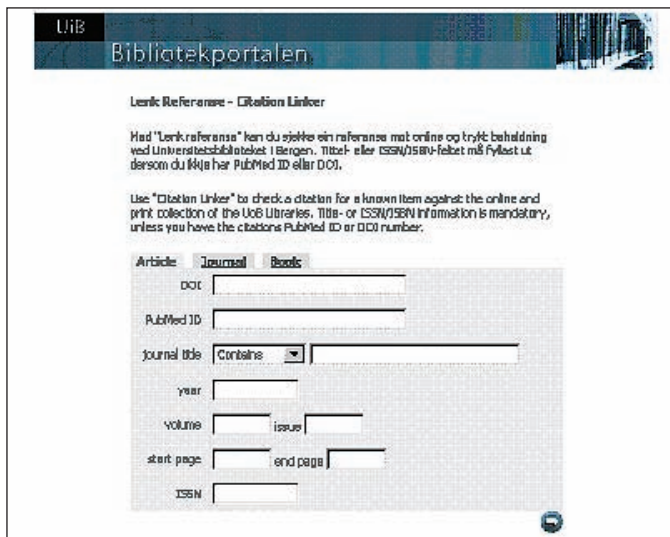
Before proceeding to log in a brief description of the rest of the interface is in order.



The portal has six basic features:

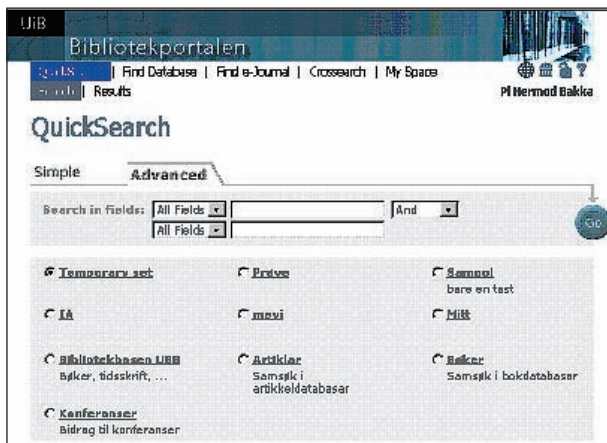
- I. The Citation linker is a “document finder” for locating verified documents using regular metadata in the SFX interface
- II. QuickSearch enables the user to (cross)search in library predefined database sets
- III. Find database enables the user to browse and search in the databases of her choice
- IV. Find e-journal enables browsing and searching for e-journals
- V. Crosssearch allows the user to search the any chosen set of databases
- VI. My space is where the logged-in user stores the databases and/or e-journals he feels he has regular use for.

I. The citation linker is, paradoxically not part of the “Portal”. It is a user interface for the SFX-facility and looks like this:

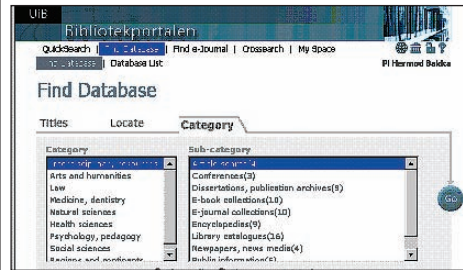
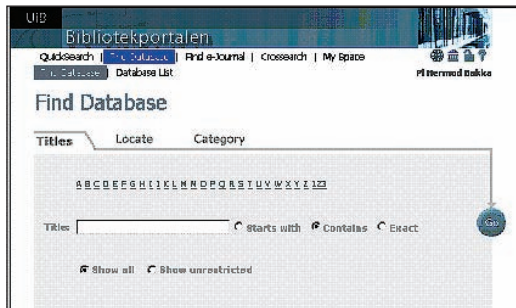


The most important added functionality of a portal compared to the catalogue is the ease with which electronic journals can be accessed. The open URL protocol allows a user to enter a normal reference into the form and go directly to the document itself. While journals and books can be searched for, it is primarily an “electronic article finder”.

II. The **Quicksearch** screen for a logged-in user may look like this:



The author’s interface contains his personal database sets (upper two lines) and the sets predefined for all users. As the sets shown are linked to the user profile this facility enables the library to show predefined database sets based on the known needs of specified user groups: i.e. bachelors’ or masters’ or PhD students.



III. The **Find Database** screen is the default access interface. It allows for browsing by database title or by library defined category.




The content of the portal may be arranged in any way the library desires. The University of Bergen Library has chosen to organize its categories in a hybrid and eclectic manner –but one which seems to satisfy the needs of our users. One important consideration is simply that the number of resources in any one category ought not to be too high. Once the number of entries in a list exceeds that which can be shown in a single full screen, portal functionality is degraded.

Categories can be divided into sub-categories such as the interdisciplinary “librarian” categories shown. Again, sub-categories can be arranged according to the needs of the library and its patrons. The Social Science category is thus arranged both by department and subject. The list for the sub-category Comparative politics at the time of writing looks like this. As the screenshot shows it is already too long.

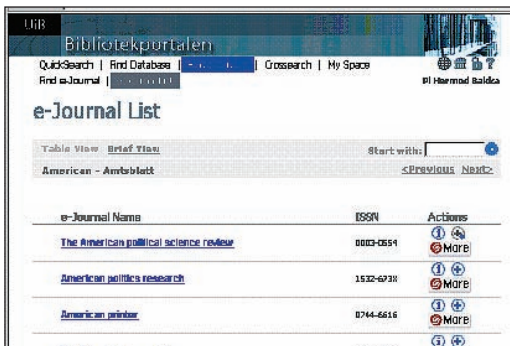
| Database Name  | Type            | Actions |
|--|-----------------|---------|
| <a href="#">African American Studies - JSTOR</a> <b>JSTOR</b>                      | E-Journal       | 🔍 📄 🔄   |
| <a href="#">African Studies - JSTOR</a> <b>JSTOR</b>                               | E-Journal       | 🔍 📄 🔄   |
| <a href="#">Asian Studies - JSTOR</a> <b>JSTOR</b>                                 | E-Journal       | 🔍 📄 🔄   |
| <a href="#">BIBSYS biblioteksystemet</a>   | Library         | 🔍 📄 🔄   |
| <a href="#">BIBSYS biblioteksystemet</a>   | Library         | 🔍 📄 🔄   |
| <a href="#">EconLit (ESP)</a>  | Index           | 🔍 📄 🔄   |
| <a href="#">Economist Intelligence Unit (EIU)</a>                                  | Subject Gateway | 🔍 📄 🔄   |
| <a href="#">Europa World Yearbook Online</a>                                       | Subject Gateway | 🔍 📄 🔄   |
| <a href="#">Global Development Finance</a>   | Index           | 🔍 📄 🔄   |
| <a href="#">History of Science - JSTOR</a> <b>JSTOR</b>                            | E-Journal       | 🔍 📄 🔄   |
| <a href="#">IBSS - International Bibliography of the Social Sciences</a>           | Index           | 🔍 📄 🔄   |
| <a href="#">Institutt for Samfunnsforskning</a>                                    | Subject Gateway | 🔍 📄 🔄   |
| <a href="#">International Encyclopedia of the Social &amp; Behavioral Sciences</a> | Encyclopedia    | 🔍 📄 🔄   |
| <a href="#">ISWeb of Science</a>   | Index           | 🔍 📄 🔄   |
| <a href="#">Journals</a>   | Subject Gateway | 🔍 📄 🔄   |
| <a href="#">Latin American Studies - JSTOR</a> <b>JSTOR</b>                        | E-Journal       | 🔍 📄 🔄   |
| <a href="#">Middle East Studies - JSTOR</a> <b>JSTOR</b>                           | E-Journal       | 🔍 📄 🔄   |
| <a href="#">MUIP - Moroccan Studies Institute</a>                                  | Subject Gateway | 🔍 📄 🔄   |
| <a href="#">Oslo</a>   | Subject Gateway | 🔍 📄 🔄   |
| <a href="#">Political Science - JSTOR</a> <b>JSTOR</b>                             | E-Journal       | 🔍 📄 🔄   |
| <a href="#">Politics and International relations</a>                               | Subject Gateway | 🔍 📄 🔄   |
| <a href="#">Politics Science</a>   | Index           | 🔍 📄 🔄   |

The brief view presents journal name in the form of a stable URL, the ISSN number and three action options.

Clicking the title brings the user directly to the journal homepage. The type indicates the nature of the database.


The actions icons provide information  about the database, allows for saving  it in My Databases, while the presence of a magnifying glass  indicates that the database can be searched in the portal interface. This is important as only such databases can be crosssearched.

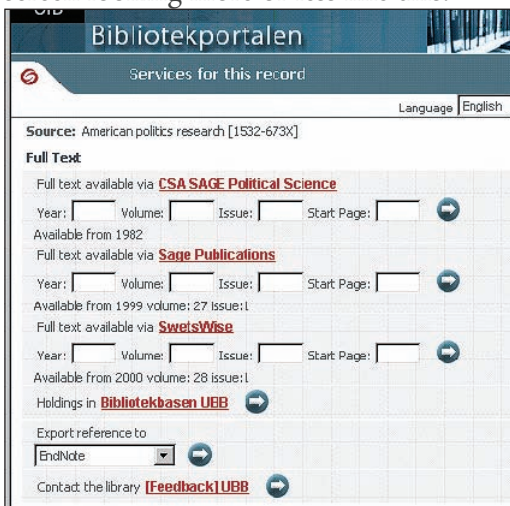
IV. At the time of writing the **Find E-journals** interface only allows alphabetical searching and browsing by title.



The screenshot presents a search result. The brief view presents journal name in the form of a stable URL, the ISSN number and three action options.

Clicking the title brings the user directly to the journal homepage.

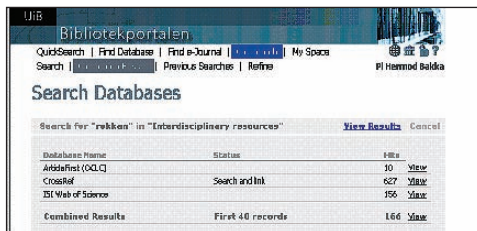
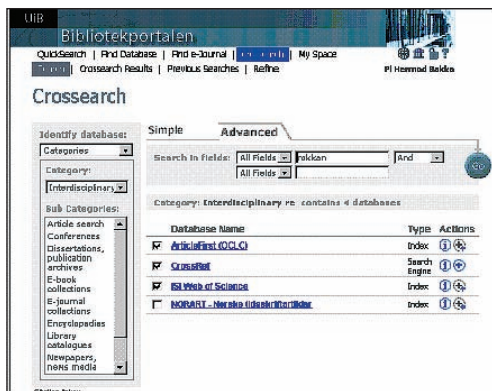
In this list the  is replaced by the key feature of the portal, the SFX-button "More". Clicking this button anywhere on the internet will bring up a screen looking more or less like this:



The SFX-screen gives the user three choices:

Filling in the form with the correct article reference and going directly to the article, searching the local OPAC or exporting the reference to a reference management system.

V. It is the **Crosssearch** functionality that brings out the innovative power of the portal. The default crosssearch screen in Bibliotekportalen invites users to perform an article search in four major databases.

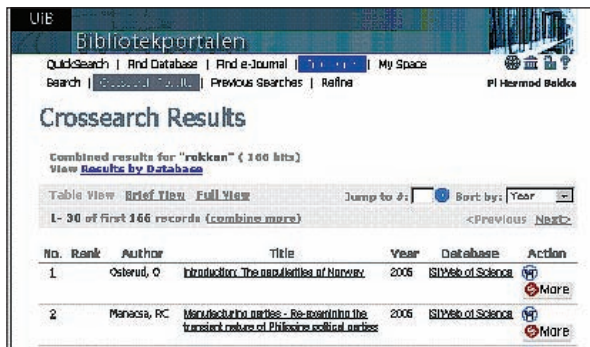


But any of the preconfigured databases can be crosssearched, as can any personal selection of databases a user defines as a database set.

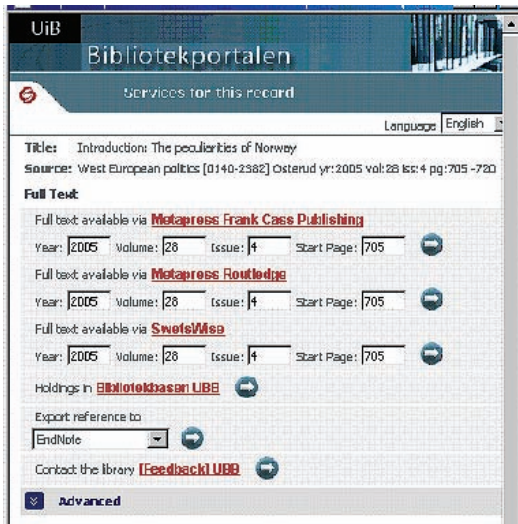
The search shown is for the term Rokkan anywhere in the content of the three databases. Two, OCLC and the Web of Science, are metadata bases, the third is the CrossRef Google full-text article search portal.

Of the 793 results 627 were found In CrossRef, a Web resource. While it can be searched using the portal interface, the results have to be viewed in the native, Web, interface.

Viewing and combining the results in the portal interface produced this list:



Clicking the SFX-button of the first entry on the list produces the screen below:



The full references for the article have been filled in and the contents are only one click away. The three separate but identical entries are due to the fact that the Bergen University Bergen Library has electronic access to the periodical in question through three different corporate channels of distribution. And that Routledge bought Frank Cass in 2003.

Clicking the blue arrow to the right of the top entry will produce this result: As each e-journal platform has its unique presentation system, getting to the text may be confusing. In the case of Metapress text is access by clicking in somewhere in the bottom right corner.

Returning to the search result it was observed that the results of the Crossref search must be displayed in the native interface; which looks like this:



Clicking the link to the article in Blackwell Synergy leads to the appropriate page in the Scandinavian Political Studies:

The screenshot shows the Blackwell Synergy interface. At the top, there are navigation links: Home, Browse, Search, My Synergy, Register, Help. Below this is a search bar and a 'Log in' button. The main content area is titled 'Scandinavian Political Studies' and features an 'Abstract' section. To the right of the abstract, there are options for 'List of Issue', 'Table of Contents', 'Prev Article', 'Next Article', 'Add to Favorite Articles', and 'E-mail this to a Friend'. Below the abstract, there is a 'References' section with a 'GO TO:' dropdown menu and a 'Go' button. The references list several articles, each with a 'More' button. The article 'Commitment and Government in Norway: Accountability in Network Politics' by Anne Lise Finnseth is highlighted. To the right of the article, there is a 'QuickSearch Inc.' section with search options for 'Synergy', 'CrossRef', and 'OR'.

The Synergy interface requires the user to access the text through clicking the PDF icon below the abstract.

It is, however, as fruitful to click the right hand, reference, button to find that the references contain SFX-buttons for immediate access to the cited articles. Even hardened librarians cannot fail to get excited over the possibilities inherent in this aspect of portal functionality.

VI. To get the most out of the Bibliotekportal it is necessary to use the **My Space** function. The default interface looks like this:

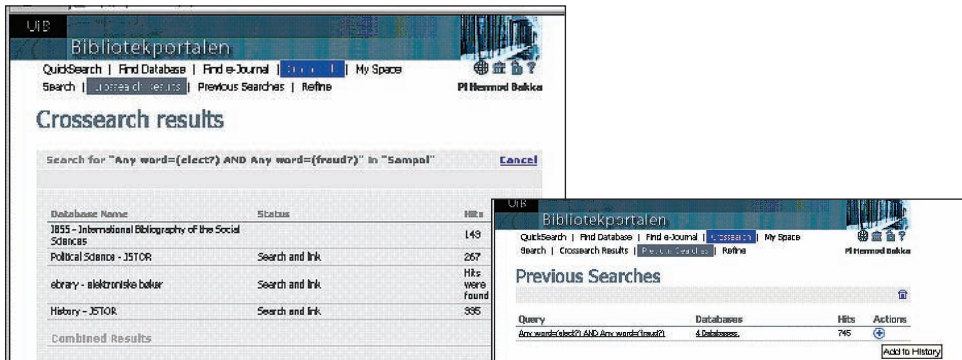
The screenshot shows the Bibliotekportalen interface. At the top, there are navigation links: QuickSearch, Find Database, Find e-Journal, My Space. Below this is a search bar and a 'Log in' button. The main content area is titled 'Crossearch' and features a search interface with 'Simple' and 'Advanced' tabs. The search fields are 'Search in fields: All Fields' and 'Search: [ ]' and 'And: [ ]'. Below the search fields, there is a table of databases with columns 'Database Name', 'Type', and 'Actions'. The table lists several databases, including 'ISS - International Bibliography of the Social Sciences', 'Political Science - JSTOR', 'Library - elektroniske bøger', and 'History - JSTOR'. Each database has a 'More' button and a 'PDF' icon.

The eShelf is used for storing electronic documents, My Databases and My e-Journals are for the preferred and frequently used databases and journals, History

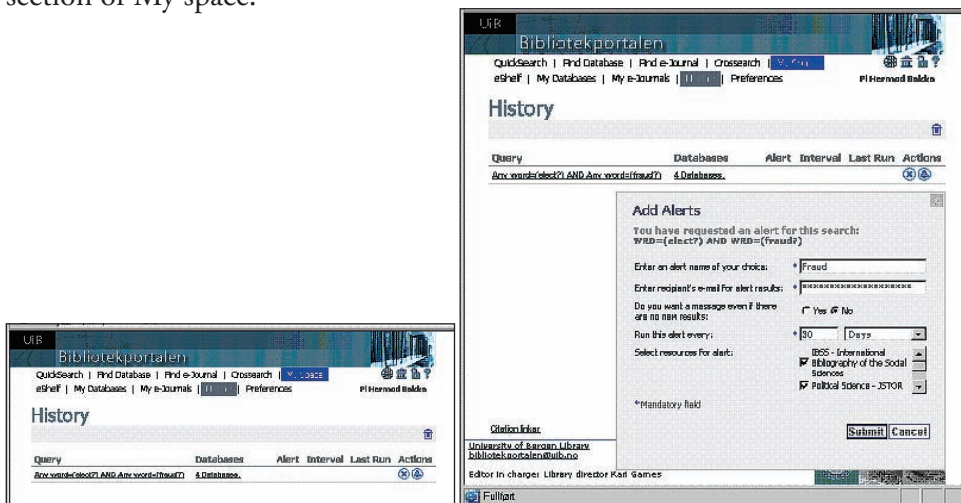


is where very large or permanent searches are stored while Preferences allow for the modification of the user profile.

Crosssearching my database Sampol for literature on “electoral fraud” or “fraudulent elections” in different types of resources I get far more hits than I can explore in session.



Accessing the Previous search function the search can be stored in the History section of My space.

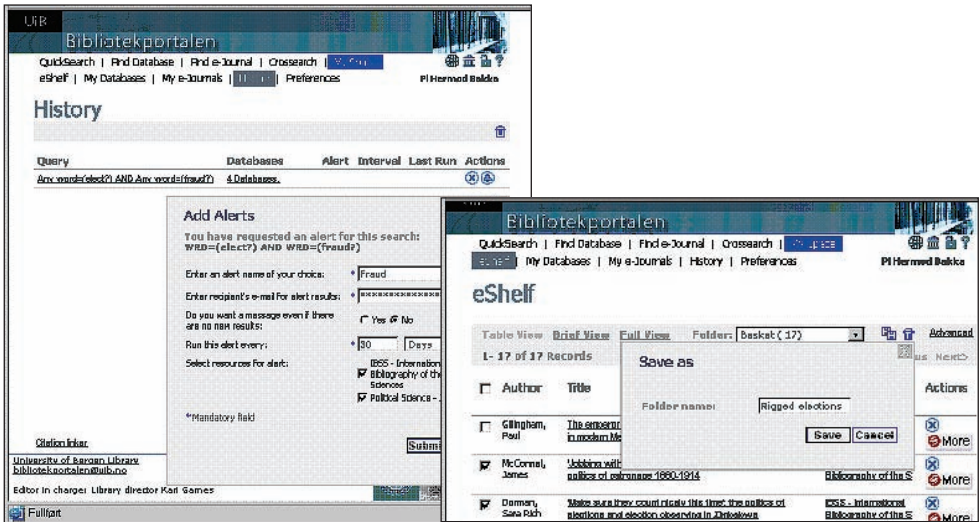


The search can be frozen or made into a permanent search clicking the Alert function (the bell icon).

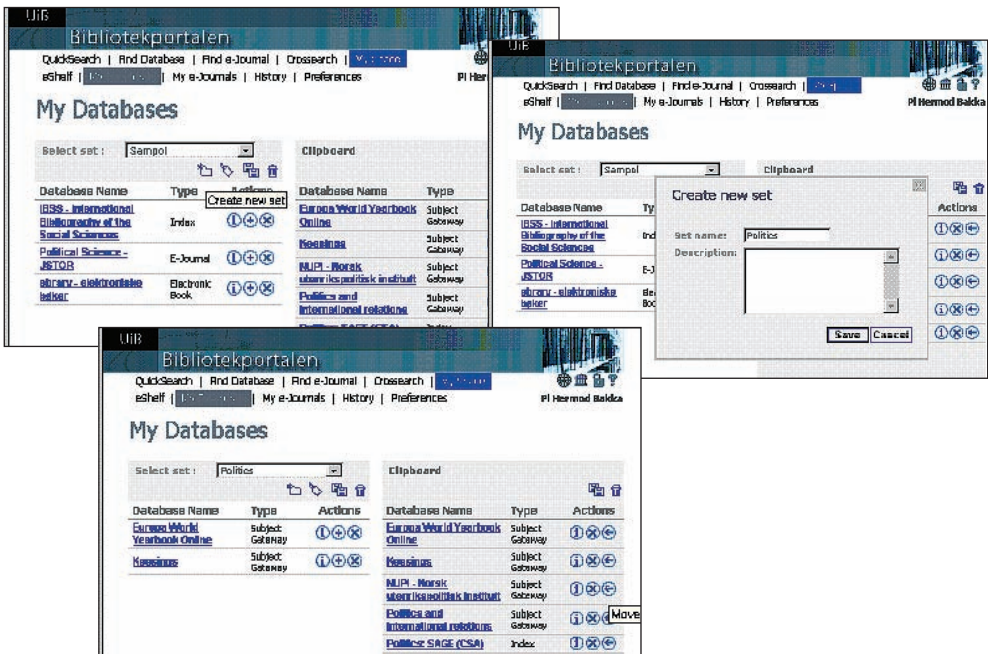
This opens the screen below, allowing for the naming the permanent search and specifying its frequency.

Clicking the link reactivates the search.

Opening the results screen allows for the references found to be put in the basket and transferred to eShelf for further treatment.

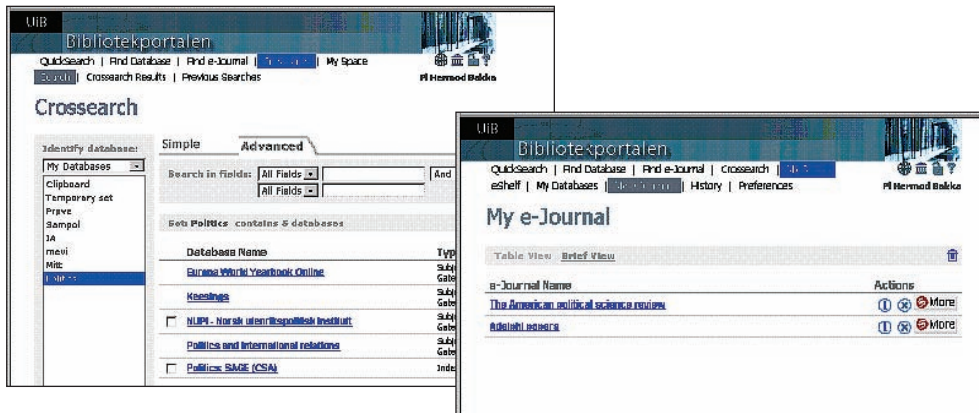


Items found in the search for electoral fraud can be saved in a separate folder, in this case “Rigged elections”. Any number of references can be stored.



The My Database facility allows users to create their personal QuickSearches in their preferred databases. Databases found can be moved to a clipboard in My Databases and stored as Quicksets.

The user may only create 13 sets on his shelf. While any limitations are unfortunate, 13 sets are probably sufficient for most users.



New sets can be Crossearched, subject to the limitations imposed by the possible lack of compatibility between the portal and native interfaces.

My e-Journals are created in exactly the same way as My Databases.

There are regrettably still some limits to what can be done in the *Bibliotekportalen*. The interface problems have been noted. The most irritating one is that individual e-Books from the major commercial suppliers cannot be stored on My Shelf. Individual e-journals cannot be crosssearched with databases. But these are all **technical** limitations. Such limits exist to be overcome.

## Challenges and Possibilities.

At the time of writing *Bibliotekportalen* has been operational for 10 months. It works far better than Bibsys did at the same point of time in its implementation 22 years ago. A library is a work in process. It is not a project. It is never “finished”. Content will be continually added. New functions will be added as well. The current technical and content limitations should not blind us to its potential as our means of organizing information for our users in a rational, easily accessible and cost-efficient manner.

The near term challenge is of course to educate our patrons into using the portal. Not by preaching, but by example. Like library and information competence education in general this is no doubt an uphill struggle. While SFX is much appreciated, faculty still relies on their personal link collections for accessing electronic periodicals. It is equally hard to wean students away from Google. But it took time to wean them away from the card catalogue also.

For *Bibliotekportalen* to become the information tool of choice for the University of Bergen community three hurdles must be passed.

The *Bibliotekportal* must become more visible on the Web, more closely integrated into the faculty Intranet and the Student Portal.

A real single sign-on system for all university ICT-resources must be introduced. At the moment full use of its potentials require no less than five separate sign-ons – which are four too many. Off-campus access must become easier as well.

The final hurdle is making students and faculty comfortable with the full range of electronic information sources. But their continuing love-affair with the printed book probably reflects the ambivalence we all feel towards the Brave New Digitized World.

# Institutional Repositories

On institutional repositories, how they came to be, and how they are fitting into the digital library

by Richard Jones

## Introduction by way of a brief history

The first seeds of the institutional repository can be traced back as far as the seminal articles by William Gardner and Stevan Harnad in 1990, when networked electronic communication was starting to become a viable tool for the dissemination of scholarly literature. In his article “Scholarly Skywriting and the prepublications continuum of scientific enquiry”, Harnad states that:

“The whole process of scholarly communication is currently undergoing a revolution comparable to the one occasioned by the invention of printing.”  
(Harnad, 1990)

These early moves towards digital repositories which really took off some ten years later were primarily suggestive of disciplinary archives, borne out by the establishment in 1991 of the Los Alamos arXiv<sup>1</sup> for high-energy physics (now based at Cornell). The first scholarly recorded proposal for an institutional archive came later, in 1994, in response to Harnad’s “Subversive Proposal for Electronic Publishing” (Okerson and O’Donnell, 1995), in which he states:

“For centuries, it was only out of reluctant necessity that authors of esoteric publications entered into the Faustian bargain of allowing a price-tag to be erected as a barrier between their work and its (tiny) intended readership, for that was the only way they could make their work public at all during the age when paper publication ... was their only option.”

The subsequent discussion which took place by email on various lists (particularly Virginia Polytechnic Institute’s VPIEJ-L list and the University of Vermont’s SERIALST list) among interested scholars, including Paul Ginsparg, the originator of arXiv, brought to light many considerations. These included the technical requirements at the time, centralised versus decentralised storage models, and the formulation of the copyright issues that would play a dominant role in the self-archiving debate. Here, Nobel Prize winner Joshua Lederberg introduced into the discussion the idea of institutional rather than disciplinary archives:

---

<sup>1</sup> [www.arxiv.org/](http://www.arxiv.org/)

“...instead let each institution set up its own ftp-able archives for all of its scholars. That way, each place can also set up its own ground rules.”<sup>2</sup>

It is notable that this discussion focused around setting up FTP archives for scholars to use with immediate effect. The institutional repository in its current form is a centralised service, often run by the institution’s library and the applications themselves require internet technologies to be well developed; therefore, the concept could not become as it is today without the time for the underlying technology to mature. The discussion and movement was also primarily driven by the scholars themselves rather than librarians; this is not surprising, since the practice of circulating preprints of articles between academics has long been commonplace. Again, though, the skills required to operate the modern institutional repository have long resided in the library, and their involvement is virtually necessary.

It is perhaps for these reasons that a large corpus of literature on the subject did not start to develop until around the turn of the millennium. Between 2001 and 2003 there was an explosion of articles covering the groundwork for institutional repositories, most notably “The Case for Institutional Repositories” (Crow, 2002) which came from the Scholarly Publishing and Academic Resources Coalition (SPARC), set up by the Association of Research Libraries (ARL) in 1998 to address the issue of high scientific journal prices.

In addition to this, software to support the creation of e-print archives really started to become available in 2001 with the release of EPrints.org<sup>3</sup>, and later in 2002 the release of DSpace<sup>4</sup>; between them the most dominant open source repository packages. It is worth noting that by this point the idea of archiving e-theses institutionally was already well under development, and ETD-db<sup>5</sup>, the major software package in this field, was released in 1999 by Virginia Tech and the Networked Digital Library of Theses and Dissertations (NDLTD). The role of the e-theses efforts should not be underestimated in the development of the institutional repository, since they provide the infrastructure to gather some institutional research under a degree of control by the organisation which is not so straightforward in practicing academic research.

We can suggest, therefore, that the origins of the institutional repository are fairly complex insofar as the modern interpretation of the term. Dominant factors would include:

- Pre-existing or under-development e-theses archives;
- Pre-existing departmental e-print archives

---

<sup>2</sup> The full discussion is presented abridged in Okerson and O’Donnell, 1995

<sup>3</sup> [www.eprints.org/](http://www.eprints.org/)

<sup>4</sup> [www.dspace.org/](http://www.dspace.org/)

<sup>5</sup> [scholar.lib.vt.edu/ETD-db/](http://scholar.lib.vt.edu/ETD-db/)

- Grass-roots faculty practices of making e-prints available on personal web pages (Johnson, 2002);
- Subject repositories such as arXiv;
- Institutional desire to preserve for both posterity and portfolio;
- Support for the philosophy of Open Access;
- Pre-existing distributed document servers;
- A reaction to the “Journals Crisis”.

Nearly simultaneous to these factors coming together to produce the first fully featured institutional repositories, the development of the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)<sup>6</sup> made the desire for interoperability and cross-searching disparate repositories a much more viable goal. First convened in 1999, the OAI group produced the first stable version of the protocol in 2001. By utilising simple Dublin Core<sup>7</sup> at the most basic level, and by providing a simple URL based query schema, this protocol dramatically lowered the implementation barriers for interoperable archives.

A critical factor, though, underpinning much of the development of all forms of Open Access repository has been the so-called “Journals Crisis”. ARL statistics show that serials costs between 1986 and 2004 increased in price by approximately 273%, while the Consumer Price Index (CPI) increased, approximately in line with library budgets, by around 73%<sup>8</sup>. This major disparity in the rate of increase of cost has caused numerous budgetary crises in libraries struggling to balance the need to maintain journal access with the rapidly diminishing budget available for other essentials, such as books. This problem has been further complicated by the switch to e-journals by publishers and the introduction of the “Big Deal”. This works by offering electronic journals on a multi-title, multi-annual basis, and has a relationship to the institution’s print cost such that the lower the print spend, the higher (relatively) the electronic spend. The outcome is that switching to electronic delivery only is not necessarily cheaper.

This problem is particularly relevant for institutional repositories, since the cost bearers are the institutions themselves. Therefore, while disciplinary repositories came first, institutional repositories can be seen as a way that an institution can directly address the issues themselves, with the long-term goal of reducing the e-journal costs, or even altering the publishing model entirely. The effort required, though, is not to be underestimated, and the stakes are high for the main participants (faculty, librarians and publishers), while overcoming the traditional publishing paradigm is a major challenge (Johnson, 2002).

<sup>6</sup> [www.openarchives.org/](http://www.openarchives.org/)

<sup>7</sup> A basic 15 element metadata set, often thought of as the ‘lowest common denominaor’ for metadata. See <http://www.dublincore.org/>

<sup>8</sup> [www.arl.org/stats/arlstat/graphs/2004/aexp04\\_pub.pdf](http://www.arl.org/stats/arlstat/graphs/2004/aexp04_pub.pdf)

The remainder of this chapter discusses how institutional repositories have been defined, how they are compared and contrasted with other similar kinds of repositories, and what kinds of uses they have found within their host institutions. Since the institutional repository, although old enough in concept, is still young in implementation, we will look at some examples of how they are being configured and presented, and the place that they have found themselves within the Digital Library. It is also of interest for us to delve briefly into the more technical details and concepts that underpin the repository, including common and desirable features, digital preservation activities (and especially the Open Archival Information System reference model), and information management challenges that are presented. We will also look at the social features of the repository such as how it is viewed in relation to intellectual property rights, and the challenges that libraries face in producing advocacy strategies to populate their repositories.

### **The many faces of the institutional repository**

Institutional repositories tend to have a very wide remit. They mean many different things to many different people, and are used in a variety of ways. The sorts of content types, for example, that we see include e-prints (both pre- and post-prints), gray literature (especially e-theses), working papers, technical reports, books and book chapters, conference papers and posters, and even some administrative records. Some broad working definitions have been drafted which try to encompass the functions of this repository type. For example, Clifford Lynch, the director of CNI<sup>9</sup> defines it as:

“...a set of services that a university offers to the members of its community for the management and dissemination of digital materials created by the institution and its community members. It is most essentially an organizational commitment to the stewardship of these digital materials, including long-term preservation where appropriate, as well as organizational and access or distribution” (Lynch, 2003)

While Lynch defines the institutional repository as primarily service orientated, and by commitment from the institution, Raym Crow of SPARC defines it from a slightly different, but compatible, angle in terms of the origins of the material as:

“any collection of digital material hosted, owned or controlled, or disseminated by a college or university, irrespective of purpose or provenance” (Crow, 2002)

---

<sup>9</sup> [www.cni.org/](http://www.cni.org/)



Crow goes on to note that the college or university boundary is not absolute, and that organisations finding use and benefit from an institutional repository could include government departments, non-governmental organisations (NGOs), museums, independent research organisations, federations of societies and even commercial entities.

Further to looking for broad definitions of the repository, there have also been efforts to define some general properties that either define the institutional repository, or are the natural outcome of maintaining one. From the literature available it is possible to define six characteristics without too much difficulty (Crow, 2002; Johnson, 2002; Lynch 2003; Genoni, 2004) :

- Institutionally defined;
- Scholarly;
- Cumulative and perpetual;
- Open and interoperable;
- Capturing and preserving events of campus life;
- Searchable within constraints.

We will not discuss the justifications for these characteristics in depth, but some notes on the second and third points are warranted. That a repository be *Cumulative and Perpetual* suggests that it continues to gather materials, and continues to store those materials (under some selective preservation strategy) in perpetuity. That is, the archive is not static and it is not short-term. Meanwhile, being *Open and Interoperable* is the only way that an institutional repository can reasonably interact with other repositories, and placing access barriers of any kind will no doubt lead to a lack of use globally, and ultimately obsolescence (which would make being *perpetual* considerably more challenging). There are further discussions on these characteristics in “The Case for Institutional Repositories” (Crow, 2002) and “The Institutional Repository” (Jones et al, 2006:53-54).

We could try to contrast these properties with other library information systems such as subject repositories, learning object repositories, institutional record archives, library catalogues and metadata aggregators. The boundaries between all these are extremely blurred, and we find, for example, institutional repositories containing learning objects or institutional records. In addition, the term “learning object” itself has a fuzzy definition, and could encompass materials such as journal articles. Areas which make the institutional repository what it is, then, appear to be its institutional branding for the end-user, its further focus on materials that reflect the nature of institution members irrespective of their use in the outside world, and perhaps the idea that it is related to preservation more so than other repository types (in that it has a highly interested curating entity).

What we generally find is that most modern institutional repositories are primarily e-print and e-thesis archives, most likely because of the development history and origins of the repository. Other content types such as multimedia, course materials and datasets are emerging, but they are yet to be considered commonplace.

When choosing how to organise institutional repository holdings, which we expect to reflect the fact that it is institutionally defined, we find a number of things. The obvious pattern to follow is that of the institution's internal organisation, and this structure is common. We often find that at the lower levels of the categorisation that a content type (e.g. e-theses, e-prints) structure is used to further refine the collections. Some repositories are actually finding themselves cross-institutional, either because of common research goals in small organisations, convenience in terms of technical support or geographical proximity and branding.

Other collection structures that are found include those which rely purely on categorisation by content use type (e.g. technical documentation, learning objects), while some use a very pure subject heading structure. We also see that many repositories exhibit inconsistent structures, which can be ascribed to a number of causes: first, that devolved administration produces structures in one area of the repository that are not informed by those in other areas; second, that some structures may be more appropriate to different disciplines or organisational units; and third that due to the youth of many of the systems their place is uncertain in their institution and information environment.

### **The institutional repository in the digital library**

If the institutional repository does not yet inhabit a defined place in the information environment, then they are not sufficiently well established to even be considered essential elements. It may be, for example, that Open Access Journals and disciplinary repositories will prove the most effective and popular in the long term. Nonetheless, they are creating for themselves a place in the Digital Library, itself a relatively new entity. Aside from the traditional library Online Public Access Catalogue (OPAC), other components of this environment include the increasing prevalence of portals, both library, institutional and even disciplinary. Electronic journals are a staple part of the digital library environment, as well as e-books and internal services such as departmental web pages containing course and reading lists, for example. Learning object repositories are also making appearances, and there is always the wealth of general information literature that has been provided by libraries since before the electronic age containing help and tutorials for users. The rest of this book introduces many of these facets, and we can see how rich the information environment becomes.

These content types exhibit large amounts of cross-over. For example, materials held in institutional repositories could be catalogued in the OPAC, while it is often necessary to surface resources in multiple portals, each of which may have differing ideas as to implementation; repositories will also often contain material that is available in some form in an e-journal, and traditional library resources could be seen as relevant material for storage in the repository. The outcome of this is that the repository must be able to interoperate: exporting records for OPACs, providing web service interfaces for portals, and being flexible enough in metadata capture to hold a variety of unusual or unpredictable objects. Many regard the institutional repository as one kind of storage mechanism among a sea of other systems appropriate for their use type (for example, the advanced management and presentation of image collections may be better done elsewhere). That is, the institutional repository is not the complete solution for institutional digital asset management, but it may rapidly become an important part. Creating well organised networks of information will be the ultimate goal, and the repository will be one of the participants.

In addition to this, the global information network of different types of repositories is increasingly gathering more nodes. Repositories should, as Crow notes, be open and interoperable, and using technology such as OAI-PMH we are seeing the rise of repository information networks as theoretically depicted

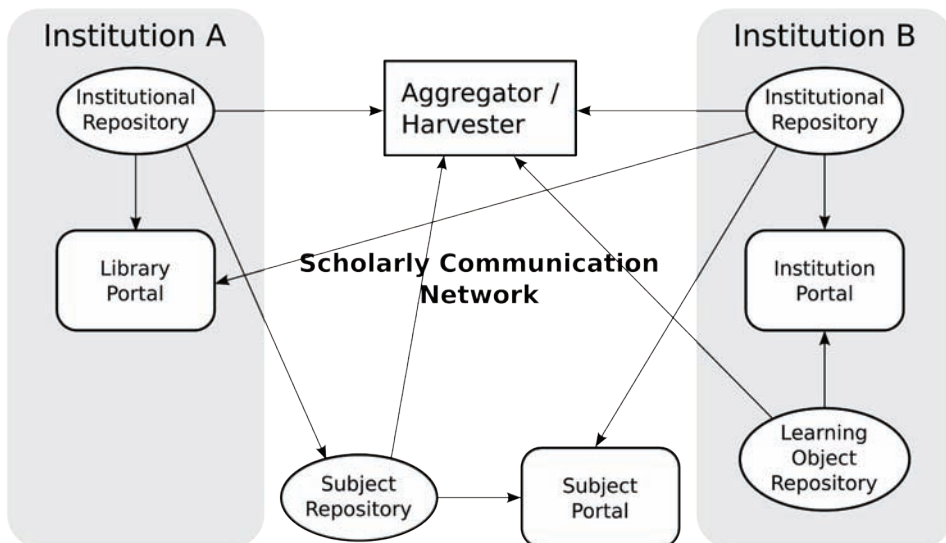


Figure 1. A theoretical repository, portal and harvester network for scholarly communication

in *Figure 1*. Here, the different information sources, ranging from institutional and disciplinary repositories through to databases of aggregated metadata from all types of repository are interacting. The goals of this interaction include making available information in as many places as possible and providing users with the opportunity to access materials in a number of different ways, and ideally in a full-text form. Cross searching, and tools such as OpenURL make tracking down the desired information that much easier, and if one route is not necessarily open to a user, another may be.

### **Technicalities and details**

It is useful, in order that we understand what the implementation of an institutional repository really brings with it, for us to examine in more detail the sorts of features and requirements that might be appropriate, as well as the complicating social issues that arise. We can start with an idealised approach to the general features that would be appropriate for a repository of this nature. Many of these systems will exhibit some of these features, although it is a list that may ultimately be attained rather than that already exists. It should also be borne in mind that institutional requirements for the repository will vary, since we have noted that it is not a well-defined system. The following, nonetheless, outlines are fairly fully functional institutional repository (Jones et al, 2006:55-62):

- *A strong development community*; particularly important in open source software, a community behind the package driving the repository is always an asset. This is an aid to technical support, further development and so forth.
- *Easily integratable*; since the repository will be one part in an already established electronic landscape it is useful to have a system which can easily embed itself both technologically and culturally.
- *Security*; some particular uses of the repository will require direct interaction with particular users, and the facilities to authenticate these users and authorise their activities is necessary. In line with integration, a system which can talk to established institutional security mechanisms is ideal.
- *Archival integrity*; to maintain holdings with some certainty over time it is necessary to ensure that content is not tampered with, and that the audit trail and provenance of the item are kept in good order.
- *Administrative tools*; repositories typically require tending by trained administrators. Good tools to aid this administration become very important as the popularity of repositories increases.
- *Licensing and Restrictions*; while we aim to capture as much material as

possible, some may still have legal barriers to open access. The repository, in these cases, should try to hold the material under a relevant licence and restrict public access until such time as it can be released as opposed to rejecting the content.

- *Web service interfaces*; in particular, support for the OAI-PMH is very important, but other protocols such as SRW/U and even z39.50 could have a place in the repository.
- *Metadata management*; it is necessary for both exposure to end users and to web services that metadata is in good shape and appropriate. Systems should gather adequate metadata both for description and access as well as ideally more technical metadata to aid preservation.
- *Federation, devolution and scalability*; as repositories grow both in holding size and usage it will be ultimately necessary to devolve features such as administration and storage in order to achieve the necessary scalability.
- *Ingest and egress routes*; it is of paramount importance to ensure that materials can both get in and out of the repository with relative ease. All barriers to ingest and egress must be kept as low as possible to make adoption more attractive, and interoperability easier.
- *Preservation activities*; the opportunity is available for institutions who wish to preserve their intellectual outputs to do so within the remit of the institutional repository. It can provide both the software tools and the cultural change which will gather output from all locations and make them available as targets for digital preservation (Wheatley, 2004).

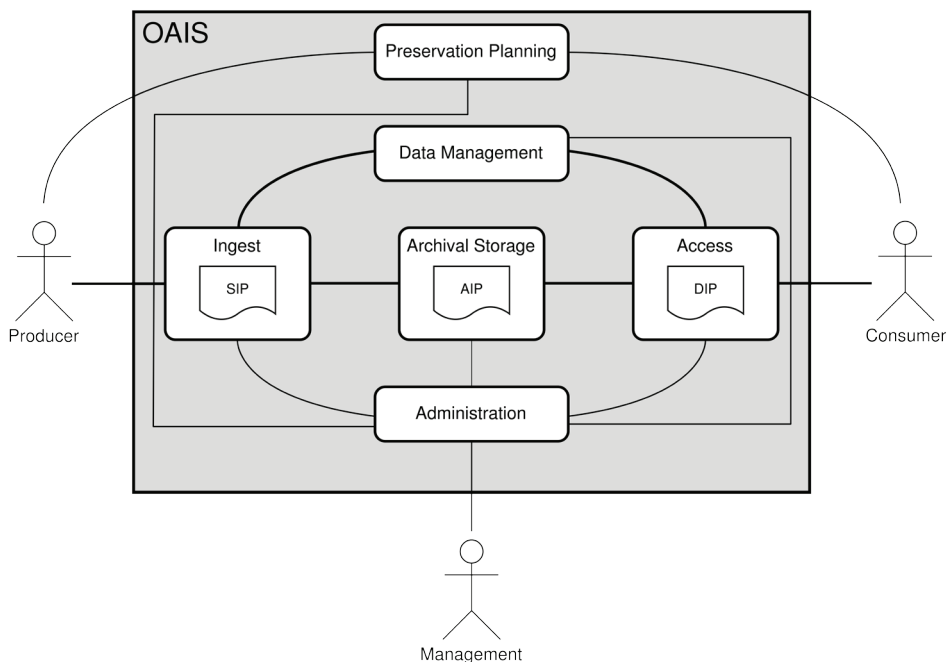
We also see repositories being imbued with other characteristics, whose value-added services could be the gateway to really embedding them within institutional working practices. These include tools such as researcher home pages, which can act as a full-text CV for academics. Full-text searching is common, and we are finding increasing use of true subject classifications, making the institutional repository more and more a core service. End user features such as annotation and discussion forums for items have appeared, and the option to have the item printed by the institution is being offered in some cases. The CDSware<sup>10</sup> team at CERN in Geneva are offering on-the-fly file format conversions to aid in digital preservation, as well as giving the option to web-cast streaming content. The repository then becomes just one facet of an increasingly sophisticated information and research tool.

The major technical challenges for the repository are really shared by many forms of modern archival systems. With the advent of grid technologies, using federated storage is becoming important both to store the digital content

<sup>10</sup> [cdsware.cern.ch/](http://cdsware.cern.ch/)

as it grows in size and to aid in preservation. Meanwhile, end user tools are constantly in development to aid cross searching of multiple databases from various institutional and subject portals, which has major challenges in terms of user interface development and interoperability standards. OAI-PMH use has become widespread, but this brings with it many information management challenges which have yet to be satisfactorily addressed; these include record deduplication (OAI, 2005) and metadata enhancement (for an example, see McClelland, 2002).

With digital artifacts as young as ten years we are already seeing a high degree of obsolescence both of format and storage media, so digital preservation is one of the big topics being addressed in the library, information and computing communities. This area has no guaranteed solutions (and may never have any), although there is a lot of development and recommendations in areas such as content migration, emulation of software environments, or even development of generic platforms upon which digital preservation activities can take place (see Wheatley, 2004 for an excellent introduction to digital preservation in the context of institutional repositories). In 2002 the OAIS Reference Model ISO standard (CCSDS, 2002) emerged as a recommendation for how an archival system can aim to support preservation activities. The basic workflow of the system is presented in *Figure 2*.



*Figure 2. OAIS reference model overview (Jones et al, 2006:78)*

*Figure 2* shows three main participants in the archival procedure: the producer, the consumer, and the manager. The producer generates a Submission Information Packet (SIP) containing all the relevant metadata and content for the item. This is then converted by the system into an Archival Information Packet (AIP), which is the preservation copy. The consumer accesses a Dissemination Information Packet (DIP) through the access area of the system, which determines which AIP to convert and deliver, based on the working data in the Data Management section. The OAIS system knows how to transform data between these forms when necessary. Across all of this, then, there is administration as handled by the manager, and preservation planning spanning all three participants. Detailed discussions on this OAIS reference model can be found in CCSDS, 2002 and Lavoie, 2004.

Some of the most critical details and technicalities of the institutional repository, particularly those that act as barriers to adoption, are social, though, not technical. Problems that were proposed early on, before many academics had an opportunity to come across the institutional repository or open access were that there may be a quality control element somehow missing from the open access repository model, and that the entrenchment of the paper publishing model, due to its monopoly, would be difficult to break.

Already Open Access Journals are addressing the issue of quality control by applying equally valid peer review. They are also trying an experiment in the publishing business model that will not necessarily mean that traditional publishers will have to fight the open access movement, but instead modify their working practices to take advantage of it. The role the institutional repository plays in this process is threefold: as part of a global effort to improve open access, they act as a catalyst for change that is directly accessible to the researcher; they can hold and disseminate materials that become available either through open access publishing or from publishers who are amenable to the self-archiving process; they smooth the path to open access by giving a forum for academics who are without an open access journal to appropriately publish their work.

Concerns that repository implementers must be prepared to deal with that stem from this slowly changing model are primarily centered around copyright and intellectual property rights (IPR). This is a clear hang-over from the entrenched publishing model, and it takes time to demonstrate that handing over copyright to a publisher is not an absolute requirement, and that the availability of materials in an open archive does not preclude immediate copyright or IPR infringement by unknown parties. The real crux of the problem is in the unclear position that repositories of all types hold within the publishing environment; contrary to traditional services, libraries are finding themselves increasingly in the role of a publisher which brings with it many new responsibilities (Jones et al, 2006:145-

147). Much material that academics would like to deposit is already copyright by the original publisher, while other material (especially theses) could be safely published online now, but damage the possibility of later publication of derived works because of prior publication. This paralysis can be worked at by chipping away at the copyright problems by asking publisher permissions to place materials online, addition by the author of friendly clauses to publisher contracts, and by encouraging authors to publish in open access journals wherever possible.

There has been much work in this area since repositories started to become reality. The Securing a Hybrid Environment for Research Preservation and Access (SHERPA)<sup>11</sup> project in the UK now maintains a list of publisher policies<sup>12</sup> with regard to self-archiving practices, and a similar service is provided by the EPrints.org community<sup>13</sup>. This allows academics to rapidly ascertain the default copyright status of their work for deposit in an institutional repository.

Repositories and their managers must mitigate their risks carefully in this field. There are large administrative overheads in ensuring that all holdings are copyright safe. While checks must be made by any responsible organisation, sometimes it is impossible to know whether an item is truly clear for deposit. The only person really in a position to know this is the author, so repositories should be sure to licence works appropriately as they are deposited. The main stakeholders in the licensing process are the author, the institution and the end-user; a licence for any deposited material should include a statement by the author that the material is theirs to deposit, preservation rights for the institution (transform and migrate and so forth), as well as reuse/distribution rights that the author can agree to in order to allow reasonable open access to their work.

One study by the Zwolle group<sup>14</sup> has attempted to identify all the major stakeholders in the creation and dissemination of scholarly materials, and investigate and promote balanced approaches to rights management. They hope that this is ...

“...a crucial step toward the development of policies or agreements that seek to assure to the stakeholders the ability to use and manage the works in fulfilment of their most important interests”. (Zwolle Group, 2001)

Further considerations in this regard by repository managers will include local factors such as Freedom of Information (FOI) legislation, as well as the necessity in some cases to implement access restrictions on holdings.

When attempting to sell the repository to faculty it is necessary to find

---

<sup>11</sup> [www.sherpa.ac.uk/](http://www.sherpa.ac.uk/)

<sup>12</sup> [www.sherpa.ac.uk/romeo.php](http://www.sherpa.ac.uk/romeo.php)

<sup>13</sup> [romeo.eprints.org/](http://romeo.eprints.org/)

<sup>14</sup> [www.surf.nl/copyright/](http://www.surf.nl/copyright/)



arguments to satisfy concerns about these legal situations. For example, a common misconception is that publishing materials in an institutional repository may increase the risk of plagiarism, and while in reality the chances appear small, and the likelihood of a plagiarist being detected are greatly increased. Repositories could also provide automated plagiarism checks as a value-added feature. For detailed information on the legal considerations for electronic resources see Oppenheim, 1999, and for further discussion of advocacy strategies see Jones et al 2006 (pp 111-138).

### **Case Study: Bergen Open Research Archive (BORA)**

The Bergen Open Research Archive (BORA) is the institutional repository of the University of Bergen. It was first launched in late 2004, as one of the first institutional repositories based on DSpace in Scandinavia. The administrative and technical base of this repository lies in the library's acquisitions department, where the skills concerning management of other e-resources such as e-journals, e-books and the library catalogue reside.

BORA is intended to work alongside other repositories managed by the library (ones which are not necessarily open to the public), such that as a whole they meet Lynch's criteria of offering services for management and dissemination of institutional digital assets. BORA then provides the public face of the university's archived research, containing research articles, working papers, books (such as this) and some masters and doctoral theses. It also broadly meets the six general criteria of an institutional repository:

- Institutionally defined: it is branded by the university and the collections are structured according to the institution's organisation.
- Scholarly: there is a wide range of purely scholarly material held therein.
- Cumulative and perpetual: it is constantly and increasingly gathering new materials, and although copyright issues have caused items to be removed in the past, most items remain and are persistently identifiable.
- Open and interoperable: it is compliant with the OAI-PMH 2.0 protocol, and provides data to the Norwegian Open Research Archives (NORA)<sup>15</sup> aggregator.
- Capturing and preserving events of campus life: it will accept most documents that are of relevance to the university, although less obvious uses have yet to be thoroughly investigated.
- Searchable within constraints: it offers a full-text search interface as well as browsing by some metadata elements including a controlled vocabulary of terms.

---

<sup>15</sup> [www.ub.uio.no/nora/](http://www.ub.uio.no/nora/)

As time goes on BORA finds itself involved in more interactions with the existing digital library environment, including providing and being a target of SFX services, being cross-searchable from the library portal, being a data-provider for several OAI harvesters, and interacting with research reporting systems. The BORA brand has also grown in the Bergen area such that it has become an umbrella term for several institutional repositories sharing a common interface, of which the University of Bergen is now only one facet.

## Conclusions

Current adoption levels of institutional repositories are pleasingly high. There are two major registries of open access archives: OpenDOAR<sup>16</sup> (Directory of Open Access Repositories), and ROAR<sup>17</sup> (Registry of Open Access Repositories). These registries show that adoption in developed countries is already very high, with some developing countries also starting to take the advantages that open access confers. The area is in rapid development, and ROAR statistics show a steep incline from around 2001 to the present in the availability of both open archives and open records in those archives available via OAI<sup>18</sup>. By simultaneous reference to the increase in literature which follows a similar timeline, we can surmise that we will continue to see further growth in this area in the short term at least.

A successful institutional repository requires institutional commitment; startup resources are relatively high, and there is a necessity for both technical and advocacy skills to be employed. To make the repository part of every day working practice of the academic is no small task, and tireless commitment to marketing and improving the service, as well as reacting to feedback from the users is necessary.

The true maturity of the institutional repository is some way off, and it will not reach it until the many copyright and working practice obstacles have been successfully tackled. Technologically, the existence of the institutional repository is straightforward, although we will continue to see new technologies, especially in the field of federated storage and devolvable interfaces, being incorporated into repository software.

This chapter has attempted to expose the institutional repository in the place that it has started to find for itself within the digital library. This landscape is constantly shifting as new technologies come and go, and the repository, like many of its information system peers, will have to adapt to changes rapidly, whilst

---

<sup>16</sup> [www.opendoar.org/](http://www.opendoar.org/)

<sup>17</sup> [archives.eprints.org/](http://archives.eprints.org/)

<sup>18</sup> These statistics only cover open archives registered with ROAR and which support and make available records using OAI-PMH. See <http://archives.eprints.org/index.php?action=analysis>

always keeping in focus its original goals and intentions. With the pervasiveness of the internet, and the large number of high quality public search services, repositories like this are going to need to form the cornerstone of the institution's commitment to bringing quality information into the public domain; a challenge which they should take to with great determination.

## **Bibliography**

- CCSDS. 2002. CCSDS 650.0-B-1: *Reference model for an open archival information system (OAIS)*. Blue Book 1. Available from: <http://ssdoo.gsfc.nasa.gov/nost/wwwclassic/documents/pdf/CCSDS-650.0-B-1.pdf> [accessed: 9 Feb 2006]
- Crow, R. 2002. *The Case for Institutional Repositories: A SPARC Position Paper* [online]. SPARC. Available from: <http://www.arl.org/sparc/IR/ir.html> [accessed 9 Feb 2006]
- Gardner, W. 1990. *The Electronic Archive: Scientific Publishing for the 90s*. Psychological Science 1 (6): 333-341.
- Genoni, P. 2004. *Content in Institutional Repositories*. Library Management 25 (6-7), pp300-6.
- Harnad, S. 1990. *Scholarly Skywriting and the Prepublication Continuum of Scientific Inquiry*. Psychological Science 1 (6): 342-344.
- Johnson, R. 2002. *Institutional Repositories: Partnering with Faculty to Enhance Scholarly Communication* [online]. D-Lib Magazine, 8(11). Available from: <http://www.dlib.org/dlib/november02/johnson/11johnson.html> [accessed 9 Feb 2006]
- Jones, R., Andrew, T., MacColl, J. 2006. *The Institutional Repository*. Chandos Publishing, Oxford.
- Lavoie, B. 2004. *The Open Archival Information System Reference Model: Introductory Guide* [online]. Digital Preservation Coalition. Available from: [http://www.dpconline.org/docs/lavoie\\_OAIS.pdf](http://www.dpconline.org/docs/lavoie_OAIS.pdf) [accessed 9 Feb 2006]
- Lynch, C. 2003. *Institutional repositories: essential infrastructure for scholarship in the digital age*. ARL. Available from: <http://www.arl.org/newsltr/226/ir.html> [accessed 9 Feb 2006]
- McClelland, M., McArthur, D., Giersch, S., Geisler, G. 2002. *Challenges for Service Providers When Importing Metadata in Digital Libraries* [online]. D-Lib Magazine, 8(4). Available from: <http://www.dlib.org/dlib/april02/mcclelland/04mcclelland.html> [accessed 9 Feb 2006]

- OAI. 2005. *Implementation Guidelines for the Open Archives Initiative Protocol for Metadata Harvesting* [online]. Open Archives Initiative. Available from: <http://www.openarchives.org/OAI/2.0/guidelines.htm> [accessed 9 Feb 2006]
- Okerson, A., O'Donnel, J. 1995. *Scholarly Journals at the Crossroads: A subversive Proposal for Electronic Publishing* [online]. ARL. Available from: <http://www.arl.org/scomm/subversive/> [accessed 9 Feb 2006]
- Oppenheim, C. 1999. *The Legal and Regulatory Environment for Electronic Information* 3<sup>rd</sup> Edition. Tetbury: infonortics
- Wheatley, P. 2004. *Institutional Repositories in the context of Digital Preservation* [online]. Digital Preservation Coalition. Available from: <http://www.dpconline.org/docs/DPCTWf4word.pdf> [accessed 9 Feb 2006]
- Zwolle Group. 2001. *Copyright Management for Scholarship* [online]. SURF. Available from: <http://www.surf.nl/copyright/> [accessed 17 Feb 2006]

# **Parchment and paper in digital University Libraries – new contexts for manuscript and archival collections**

By Rune Kyrkjebø

## **Introduction**

The advent of digital editing and publishing does not mean the end has come for the Gutenberg era. Most of what we read, we still read in print, even if we locate it and select it on the computer screen. The proportion of digital library services compared with traditional book and journal ones is obviously growing at a high rate. Digital information is a rich and wonderful source, with no physical limits to it, for those who are online. For many, the internet is already the first place to turn when searching for information. In the case of scientific journals, it seems that the electronic medium is already coming close to replacing the printed one. While the situation is not the same as far as books are concerned, one could probably say that in theory at least, the electronic medium can completely replace the modern printed book – or more precisely: Technology enables books to exist primarily in digital form as structured electronic texts, which then can be formatted and printed on demand.

The impact of the digital age is not of the same type or at the same scale when we turn to manuscript and archival materials. The drive towards digitisation is naturally much stronger in the case of structured, synthesised textual information like books and journals. Those texts are ready for reading. In manuscript collections, there is typically, apart from the highlights we all want to see, large amounts of materials that contain raw, unprocessed textual information. This is not synthesised information, it is rather textual data for research and investigation. This kind of materials moves much more slowly into the digital age. The sheer size of it makes it safe to say that our paper and parchment collections are going to have to stay with us, in their original physical form, for as far into the future as we can see.

The great and potentially very rewarding challenge is to integrate manuscript and archival collections in the modern concept of a university library. In my article, I will give some principal reflections on this challenge. After saying something about the type of materials itself and its relevance, I will discuss the difference between “provenance” archives and “collections”, which I think is an

important difference to bear in mind, especially for university libraries. Then, I will move onto the role of university libraries and our physical and digital archives. Finally I point at the promising possibilities in the new educational context of university libraries involving learning centres and teaching.

My general perspective is that of a special collection department in the library of a medium size, nearly 200 years old European academic institution.

### **The relevance of unique or rare materials**

Unique materials like handwritten documents, or very rare materials like books from the earlier years of the printing age, are sometimes aesthetically beautiful objects which possess a certain aura of age and authenticity that enthrall us when we look at them. This effect is commonly experienced among the public, whenever old and rare written or printed materials are exhibited. One important duty for us as keepers of old and rare materials is of course to give the public the opportunity to see and experience old books and manuscripts. Many archives and libraries hold materials that give great opportunities for people to see the longer lines of history stretching backwards from their own spot in time and space, like the history of their country, region or city, or the history of their family, whose ancestors may be mentioned in the archives. This is an important aspect where the public justly expects our institutions to provide both access to information and context for understanding. With the electronic age, where we can scan images, digitise text and produce web publications, we have of course much better possibilities to meet this expectation. There is today an emerging wealth of high quality digital manuscript and old book publications on the internet. The possibilities to locate and access reproductions of historical materials have improved considerably over the only 10-15 years since the internet became a reality. One might wonder, in passing, what the impact of the digital world on the inherent power of the real, physical objects will be. Digital copies of manuscripts can be more colourful, larger, clearer than the original. Will the paper objects in the future become more or less powerful in their authenticity?

The relevance of the unique materials lies of course first and foremost in its provision of textual information that is nowhere else to be gained. Moreover, the function of unique textual materials as documentation of historical events and processes is of primary importance. The documentation aspect is in the archival world rightly seen as being of relevance to questions even of human rights and democracy. The documentation aspect is a major one whenever history is investigated and written, because the unique, actual and authentic materials with its information are there to demand our historical description and explanation. This goes for the history of society in general, and also, of course, for the history

and development of the academic world, where university libraries have a particular responsibility and role to play.

### **University libraries in the archival world**

As keepers of unique and rare materials, university libraries are part of a larger landscape of archival institutions. In the field of old and rare materials there is an overlap between university libraries, state and national archives, county archives, municipal or city archives and libraries, not to forget the museum sector. From the point of view of society in general, it is of little importance whether this or that material is deposited in a university library or in the state archives. The overall important points are safe preservation of materials and physical and digital, and intellectual, access for the public.<sup>1</sup> Archival institutions and libraries share to a great extent the view that the main objectives on behalf of our public audience are more important than the actual location of materials, even if each institution naturally feels privileged to have the collection that it has, no matter if the collection is large or small in numbers.

Even if archival institutions inside and outside the academic world share the main objectives and intentions regarding our duties and activities, there are also noteworthy differences in archival traditions. The processing of handwritten materials in university libraries throughout time tends to reflect a more academic perspective than is sometimes the case in the tradition of the non university archives, like state and national archives. University archives have had, in many cases at least, better time and opportunity to catalogue unique materials in detail. As far as Norway is concerned, we could say that the university catalogue tradition is more of a philological kind, sometimes richer in context and detail and more individually adapted to each archival object, than the catalogue tradition of national and state archives. If this is so, it is no wonder. It is not possible for the large receiving institutions to catalogue in deep and broad detail each document, or each series of documents, because the amounts of archival materials that each year are deposited is very large. This has been the situation for many archival institutions for a long time.

Part of the answer to this challenge is to stick strictly to the so called *principle of provenance*. This means in as much as possible to keep archival items together, complete and in the order they had when created. The fundamental value ascribed to the creator and the creation process becomes evident when one reads how the International Council on Archives (ICA) defines the concept of an archival *fonds*, which is

---

<sup>1</sup> Buckland 1991, see Skagen and Torras, this volume.

[the] whole of the records, regardless of form or medium, organically created and/or accumulated and used by a particular person, family, or corporate body in the course of that creator's activities and functions.<sup>2</sup>

One could say that the principle of provenance, when observed, keeps the materials as authentic as possible. Further, this principle preserves the order and system of the archival materials as an additional source of information about its creation. The principle of provenance is a sensible way to guard the future information value of archival materials, and it is a cornerstone of archival methodology.

One could argue, however, that the provenance principle is indispensable only when detailed, philological cataloguing is not an alternative. In principle, a detailed description of creation, acquisition, and later custodial history of the materials could preserve the same information as is taken care of by keeping the materials in the original order.

An example of a collection built around a part of an archive is the Bergen University Library MS 2053, letters to *and from* poet Olav H. Hauge. After poet Olav H. Hauge, born 1908, died in 1994, his letters were processed in the following way. All letters of strictly private character were sifted out and remained in the possession of the family. All letters concerning his activity as a poet were registered and taken over by The University of Bergen. Then, a university professor contacted the persons that had written letters *to* Olav H. Hauge, and asked them to provide the University with the letters they had received *from* the poet. A good deal of new material was provided by the recipients. Most scholars would agree that the scientific value of the collection was raised by the addition of letters *from* Hauge. From a strict provenance point of view, however, the materials have been interfered with, it *is not* the archive of the original creator anymore.

The provenance question seems to be a point where cataloguing traditions in academic institutions differ from the traditions of other important archival institutions. We do not focus on provenance in the same way as state or national archives. Some tension sometimes follows from this fact, like in Norway, where the national archives might think that university libraries are not handling archival materials correctly. Some would argue that if the principle of provenance is not strictly adhered to, we are not speaking of *archives* proper, but of *collections*. Collections, then, are implied to be of lesser archival value. The ICA defines a collection as

an artificial assemblage of documents accumulated on the basis of some common characteristic without regard to the provenance of those documents  
Not to be confused with an archival fonds.<sup>3</sup>

<sup>2</sup> ISAD(G), p. 10. See also Lange et al., p. 135 ff.

<sup>3</sup> ISAD(G), p. 10.



The question of archives as opposed to collections in my view represent a basic methodological point where university libraries have a different archival tradition. We should develop and strengthen our own tradition into the age of the digital library. I think we have to do so very consciously, since our institutions are after all comparatively small in the archival world.

It would be to oversimplify to state that philological cataloguing requires much more time and effort per shelf meter of catalogued materials, although in one sense this is true. The determining factor as to where to spend the cataloguing resources is the scholarly interest. We ask what can be scientifically gained by describing the particular archival object in detail. When scientists predict there is something substantial to gain, then there is reason for more detailed description. This we might call the *criterion of scientific interest*. The same criterion applies to the choice of materials for digitisation. Obvious as this may seem, the criterion is necessary for university libraries to make explicit to explain their aim and their job, because in it lies the cornerstone of *our* archival tradition. Universities are not here simply to collect information. Facilitating the creation of new knowledge is the ultimate goal of our libraries, with all their branches. Following this objective, university library special collections have acquired and established archival objects that are not of the “pure” provenance kind, but of the collection kind. We have good and legitimate reasons for doing so, on the basis of scientific interest.

### **The present and future role of University Libraries**

As mentioned above, the actual physical location of an archival object is of little matter to the public or even to scientists, when the object is digitally available. From the point of view of an institution, we need to ask ourselves what the collection profile should be for our physical collections as well as for our electronic ones. The profile of the physical collection is formed to a large extent by the history of our institutions. Some subjects, some periods, some types of materials will be better represented in our collections than others. Certain scientists or certain departments or projects may have produced archival materials of particular interest, archival materials that maybe today are part of the special collections.

When planning the future of our manuscript and archival collections, it will be helpful to differentiate between types of unique materials. The traditional manuscript collection typically comprises a heterogeneous materials. Written items seen as culturally, historically or otherwise valuable have been collected and catalogued here. Some items are single pages or fragments, there are single volumes of handwritten books, like diaries. There are large or small letter collections. We might also find in the manuscript collection large series of account books

or protocols from trade companies or other private organisations, maybe the voluminous private correspondence of a famous professor, along with documents created by an important research project. This approach to manuscript and archival collection, if it is the only one we take, is probably too undifferentiated and not very well suited for the future.

I will now try to outline a proactive archival collecting policy for a university library in very general terms. In the general archival collection picture, we must include both paper and parchment archives, as well as the electronic ones. There is an important link between the two types, in that the latter is often created from the former. Digitisation of physical archives is an important challenge for research institutions. Again, the theoretical scientific point of view is crucial. Paper collections of scientific value stimulate digitisation, which then adds to the potential for knowledge creation in the scientific community.

My first point is that we should keep and continue the traditional manuscript collection for one category of items. The manuscript collection is suitable for heterogeneous, often relatively small, items. The Hauge collection mentioned above, is such an item. It is not very large (about 10 folders), it is limited in scope, and it is collected and kept by the university because of the high literary and biographical research value of its contents. The manuscript collection is the right place for items like these.

When it comes to larger archives, more diverse in content and complex in internal structure, there are reasons to handle some of them differently. It is in the case of such archives that university libraries really have to take into account the larger landscape of the archival world and reflect on our place in it. Private archives (“private” as opposed to produced by a state or government public agency) are collected in all larger archival institutions, certainly also in the state and national institutions. Examples of this category are archives of corporations, firms, organisations and individuals, or research projects. When such an archive broadly documents the activity of its creator, it ought to be handled by standard archival rules and kept and treated according to the provenance principle. Now my point is that university libraries should aim at collecting this kind of materials also. The research value might not be so concentrated in larger private archives. Neither will those archives always promise a direct scientific gain. Nevertheless, *the history of our institutions* is to some extent documented by private archives created by persons, organisations, projects and activities related to our universities. The difference, generally speaking, from the typical manuscript collection item, is that private archives throw light on processes in a broader sense, while the manuscript collection item often represents a concentrated research resource. A proactive attitude towards our institutional history would be to keep and catalogue private archives at the university. The most important objective is to

preserve the archives, of course, and state or national archives would gladly do that for us. But in order to strengthen our institutional historical identity we should collect and keep important materials at our universities, for example in university libraries. Our catalogue data on private archives should then be exported into larger catalogues maintained on national level.

### **Electronic archives: Static and dynamic**

I will now complete this picture by saying something about electronic archives. Here too it is necessary to speak of different categories. One main type of electronic archival resource for universities today is the open access publication archive.<sup>4</sup> Such electronic archives offer permanent, open storage and access for digitised materials. They will become important windows into the profile, history and activities of our institutions. Such archives will accumulate a great wealth of textual information and images. Any static materials that are digitised in a publishable form could find its place in an electronic institutional repository.

When I use the word static, I imply that there exists another main category of electronic archives. What I have in mind is digital text archives that are not frozen, but have some degree of dynamics to them. This dynamics might be in terms of added textual markup to XML texts, on the fly textual transformations in order to view different aspects of the encoded text, or some other kind of continuing alteration or shifting display of the texts, or continuing input to them or rewriting of them resulting from research activities.

An example of a dynamic text and image resource is the Bergen University Library Medieval Fragment Collection (screenshot below). This web resource makes available several types of data in an integrated manner. First, there is the electronic catalogue text. The catalogue text is XML encoded, and a display of it is generated by the web server every time an enduser requests the specific catalogue entry. The text of the catalogue data is in principle a changing one. The special collections department will update the catalogue text when new information is at hand. This web resource also comprises full text transcriptions. Not even transcriptions are static texts when old manuscript materials is concerned. In the case of texts in Old Norse language, philological editing today moves in the direction of including both facsimile transcriptions and diplomatic ones in the same single electronic text file. Then the user can choose alternative views. XML encoding of additional textual phenomena may be added. Such electronic transcriptions therefore have a growing and changing character, and should be kept “alive” at least for as long as there is work going on on them.

---

<sup>4</sup> See Jones, this volume.

## The Botulph Breviary fragments

Parchment, four fragments: a: ca 17x18 cm, b: ca 25x14 cm, c and d: ca 12x3 cm, (original size: ca 25x18 cm), England, late 13th century.

**Manuscript Identification**

**Repository:** Bergen University Library  
**Signature:** MS 1549, 1, a-d  
**Alternative name (old signature):** MS 410, 1  
**Alternative name:** Br 4 (Gjerløw)


**Introduction**

The so-called Botulph Breviary fragments is a group of four fragments from the same Breviary, written in England in the late thirteenth century. One of the fragments contains the readings six to nine for the Matins of St Botulph's day (17 June, in Scotland 25 June). Nine readings constitute the highest degree of a saint's feast. St Botulph was a popular saint in Eastern England and Scotland, and the Botulph breviary was therefore perhaps written in Eastern England. The lessons six to nine, more or less extant in our fragment a, describe Botulph's discovery of Ikanho (possibly the present Iken in Suffolk, east of Ipswich), the exorcism of the area's demons, the building of the monastery (in 654), and, finally, to the great lament of his fellow brothers, Botulph's death (in 680). Apart from the readings of St Botulph's day, the fragments contain parts of the liturgy of the feast days of the Annunciation (25 March) and of the Saints Philip and Jacob (1 May).


**Manuscript Content**

a: St Botulph's day, 17 June: The fragment contains parts of paragraphs five to nine of the nocturnal readings of Matins

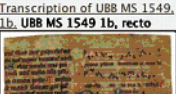
Transcription of UBB MS 1549, 1a, UBB MS 1549 1a, recto



UBB MS 1549 1a, verso.



Transcription of UBB MS 1549, 1b, UBB MS 1549 1b, recto



*Bergen University Library Medieval Fragment Collection: An electronic manuscript collection (<http://gandalf.aksis.uib.no/mpfl>).*

One characteristic of dynamic text archives like the one mentioned here, is that they require long term back up from researchers or experts in for example humanities text encoding, in addition to the service of librarians and of the central computing units at the institutions. I think that archives of this type should become a shared responsibility between faculty, library and computer department. The faculty could contribute with a long term research resource allocated to running and developing text archives. The library would be well suited to be the owner, because of its long term perspective and responsibility for all kinds of materials.

### **University libraries as a new context for archives: Learning centres and focus on information literacy**

Providing information guidance has always been an important task for special collection librarians and for archivists in other institutions alike. Knowledge about the context for the creation of manuscripts and other archival materials, and knowledge about the later history of the materials, is also something archivists have provided both researchers and the general public with. Historical and philological knowledge, information guidance and knowledge of sources, has been, and is, a strength of manuscript departments in university libraries.

Today, however, university libraries are moving on to more explicit methods and more systematic, scaled up user education. Also, cooperation between faculty and library in this respect is increasing.<sup>5</sup> Aims in university library strategies today are to establish well functioning learning centres, and to increase the information literacy of students and staff.

Special collections materials like manuscripts is not what students need to access early in their studies. At postgraduate level, however, there might be need for manuscript materials for students also. An obvious candidate field for setting up systematic teaching activity on manuscript materials is history. A university library with a collection of old documents could cooperate with the department on offering courses in reading, classifying and describing old documents. The digital era enhances our possibilities here. To study manuscripts on computer screen is often better in fact than having the real thing on your desk. Important gains for the library in connection with such activities could be improved catalogue data and the addition of electronic transcriptions of documents to our collections. This would be facilitated by dynamic electronic archives like the fragment web site presented above.

The growing teaching activity at university libraries is a fortunate circumstance for our manuscript and archival collections that allows us even more efficiently to make the most of the research and educational value of our materials. There is no doubt in my mind that university libraries are going to lead in this future development. This is a new context that strengthens the need to integrate our archival traditions with the new digital educational realities.

## **Summary**

There are certain general advantages and possibilities for university libraries with archival materials entering the age of digital media. As before, our institutions will both serve and benefit from our faculties, institutes and individual researchers and students. Being situated in the midst of knowledge creation and educational activities is a most fortunate context, both practically and strategically, when it comes to archival materials and manuscript collections. The digital aspect of libraries means there is a new common ground, or common space, for special collections and ordinary collections to enter. There we should build on and strengthen university library archival and manuscript tradition.

---

<sup>5</sup> See Vedvik Tønning, this volume.

## **Bibliography**

Bergen University Library Medieval Fragment Collection: An electronic manuscript collection

<http://gandalf.aksis.uib.no/mpf/>

Buckland, M. (1991). *Information and Information Systems*. New York, Greenwood Press.

International Council on Archives (2000). *General International Standard Archival Description*, Second Edition. Ottawa.

[http://www.ica.org/biblio/isad\\_g\\_2e.pdf](http://www.ica.org/biblio/isad_g_2e.pdf)

Lange, Vilhelm, Dag Mangset, Øyvind Ødegaard (2001). *Privatarkiver. Bevaring og tilgjengeliggjøring*. Riksarkivaren, skriftserie, nr. 11. Oslo, Kommuneforlaget.

# Digitisation as a strategy for preservation and dissemination of photographic archives

by Solveig Greve

## Background and presentation

This article introduces the Picture Collection of the University of Bergen Library (UBL), presents some reflections on the photo-conservation and preservation aspects, and describes its digitisation history. Experiences with this collection are then discussed and related to the role of digitisation in photographic collections: how, when, what and why digitise?

## The Picture collection and the Knud Knudsen Archive

The Picture Collection is a department of University of Bergen Library, the bulk of the collection consists of photographs from the 1840's to the 1970's. This historical photo-archive is among the oldest and largest in Norway. It was established in the University Library in Bergen in 1966, but an important portrait-collection had already been part of the Special Collections of the Bergen Museum Library from 1898, before the Museum became part of the new University in Bergen in 1947. The total number of units is unknown, due to a large backlog, but the active part is approximately 350,000. Although the University research- and student community use the collection, patrons external to the university more frequently use it: publishing houses, museums and the general public.

Part of the collection is the Knud Knudsen archive. Knud Knudsen was one of Norway's most distinguished pioneer photographers of the 19<sup>th</sup> century (1832–1915). The Knudsen-collection contains the photographer's own archive, approximately 40,000 negatives and prints, as well as the archive of Knud Knudsen Company that existed after his death and up until 1974. In the national preservation plan for photography (Norsk Kulturråd 1992), the Knud Knudsen-archive has the status as one of the two most important photographic collections in Norway.



Knudsen, K, 1882, *Parti af Bergen* (View of Bergen)

When the Library acquired the Knudsen-archive, it was a collection with a high degree of archival order. Knudsen was a commercial photographer who produced his landscapes and genre-pictures for the tourist market. The archive had a printed catalogue with titles and numbers, corresponding with that of the negatives and the order in the negative-cabinets. The thousands of albumen prints had been kept in a dark environment, and were in mint condition, while the physical condition of the negatives ranged from mint to very damaged. The oldest negatives were collodion glass, and some of these had bad reticulations, which would print as dark lines in the image. The most dramatic damage was to the early 1880 gelatin dry-plates; on some, the emulsion had started curling up, thus coming loose from the glass support.

The University Library bought this collection in 1975, when preservation standards were different from what they are today. The planning of future digital re-photographing as a conservation and preservation strategy started in the Picture Collection as early as 1990. Therefore, our knowledge of digital practices in photographic collections found its form through working with the conservation challenges of this collection, changing and developing in step with the technical development, the general knowledge of digital preservation of photography, and of course the available economic resources at the time.



### **Photo-preservation in Norway**

National strategies for photo-preservation started early in Norway. In 1953 the Norwegian Iconographic commission published a report on the destiny and status of the photographic heritage, called "Save the Photographs!" (Norsk Ikonografisk Kommissjon, 1953). In 1976 to 1977 the groundwork was made for a national plan of photo-preservation. A government-funded national Secretariat for Photo-registration (SFFR) was established, and a plan for a three-level organization structure was proposed: A national level, (SFFR), one institution responsible for each county (mostly museums), and the local level of archive-, library- and museum-institutions owning photo-collections. Programs for collecting, cataloguing and preserving photographs were developed by the secretariat and conferences and courses in various techniques were offered. Thus the photo-conservation institutions became a close-knit professional network. In 1997 SFFR became part of the National Museum Development, (NMU), and in 2003 the photo-conservation responsibility was placed under an Archive, Library and Museum institution, ABM, established under the Ministry of Culture.

### **Some general reflections on photo-preservation**

#### **The photograph as object and information-carrier**

Silver-based photographic objects will always be in a process of destruction and disintegration because the silver compound used in photography is chemically unstable. This is the premise of all photo-conservation. Our aim is therefore primarily to delay or stop this natural chemical process working on the original object, and secondly to re-photograph the objects to contain the information. Physical preservation is always a prerequisite for content-preservation. As obvious as this may sound, with digitisation being such a buzzword and financial support following this trend, it becomes increasingly necessary to point out that digitisation cannot solve all our problems concerning preservation of cultural heritage. This is an even greater problem concerning photographic cultural preservation. There is a strong conviction that a photograph is identical with its information content. Photography has been valued mostly as a carrier of cultural information, not being in itself a remnant of cultural history. Content-preservation has been given priority before physical conservation because as such an old original print did not have any specific value over the modern copy. This is a problem that is at the core of photography as a historical phenomenon, with serious consequences for photography preservation.

Photo-conservation is expensive if done to optimal standards. The

photo-archivist is often placed in a difficult position, finding the right balance between allocation of resources to physical preservation and to digitisation for preservation of content. Also, there is a continuous pressure from the research community and the general public of making rare and important source information easily accessible through searchable databases and internet-publication. All of this makes any digitisation project easier to finance than projects of physical conservation. Also, the fight for physical conservation often feels like a lost battle since the photographic heritage contains such masses of objects that making priorities is difficult. Photographic processes of the 20<sup>th</sup> century, (nitrate and acetate carrier and color photography) have a shorter life span than the old glass-based photographs of the 19th century. The immense photographic heritage consisting of family photography is often quite repetitive in content and therefore comparatively lower in median source value than older collections. Hence we are left with large quantities of collections of questionable value demanding a disproportional large part of our preservation resources.

### **Library practice: Public access and preservation**

The Picture collection is part of the University Library, and is run according to library-ideals; that is to make collections available to the public as far as the physical conditions permit. On the other hand, as a collection of historical artifacts, the rules for handling and loan must be different from other lending-rules of the library. In principle, re-photographing the whole collection for viewing in the reading room can solve this problem. The collection is so large, however, that this is not practical or economically feasible. We do make modern prints of the originals that are highest in demand for readingroom-viewing, but when the user needs to see images that have not been through this process, the original print must be taken out of the storage. No pictures are taken out of the library, and therefore an important part of our daily work is the production of surrogate copies from the archive for research, publication or decoration. We produce a large number of digital files, analogue prints and digital prints, dependent on the user's needs. We are operating in an environment where there is a strong demand for photographic quality, not only for reasons of preservation standards, but also from the needs of our patron researchers and publishers. A photographer specialized in working with historical photography is employed to ensure this professional standard.

Over time, the practice of re-photographing has resulted in a considerable amount of second-generation negatives of the most "popular" images, protecting the original from damage through handling. We have found this to be a more economical use of our resources, than doing en-bloc re-photographing of whole collections.

## Projects

### Project of re-photographing

After the Norwegian National Preservation plan of Photography was put into effect, the library received financial support in 1994 from the Culture Council of Norway for a content-preservation project done on analogue 6x9cm film. The National Library had opened a reservoir and technical preservation department in Mo i Rana, in the north of Norway, equipped with special cameras for doing high-end re-photography. These cameras were transportable, and thus offered picture archives the means for doing the actual photography job locally. The films were then developed and copied in Mo i Rana, where they keep the master film, sending a copy back to us. After some experimenting, most of the resulting negatives were of a high quality. However, it became clear that this was a mass-production scheme that did not take into account the variations in density, exposure and damage-profile of the old negatives. In-camera manipulation of exposure-time / aperture was possible, but controlling density and contrast through development manipulation was impossible because of the technical set-up at The National Library. A decision was therefore made to exclude the most difficult and most damaged negatives from this project for a more individual treatment.

### Project “Digiknud”

#### *Digital re-photographing for preservation*

In 1997 the University in Bergen agreed to finance equipment and a photographer for a two-year project of digital re-photography to take care of these most difficult objects and at the same time develop a standard for future digital re-photographing. It took us a long time to decide on the best equipment for our purpose. The first consideration was to choose between a scanner and a camera with a digital back. We decided on a camera for several reasons. With a camera, the size of the object to be digitized presents no problem. Also, we were afraid that the strength and quality of the light emitted by scanners would be harmful to the old photographs. Thirdly, we did not want to put glass-negatives inside a scanner, fearing they would break. The equipment decided upon was a Sinar large-format camera with a digital back, developed for medicine research and therefore having the ability to do scans with very high resolution. This project, called DIGIKNUD, ran from 1997 to 1999. Simultaneously the National Library in Mo i Rana started digitisation from the master negatives of the re-

photography project. The standard for this digitisation was set in accordance with our local tests of resolution and sharpness. The resulting files were returned to us on CD's. The digitisation standard for DIGIKNUD was set for 300 dpi and 3000 pixels on the longest side, giving files of approximately 15 – 20 Mb. This is also the standard recommended in 2005 by ABM (Oulie 2005). As a result of this collaboration, we have most of the Knud Knudsen archive re-photographed both as digital files and as analogue medium format negatives.

### *Digital restoration*

Part of project DIGIKNUD was a trial project for digital restoration of some of the damaged negatives. We tested two typical damage-types: breakages and emulsion-reticulation. The original glass-plates were printable, but handling would further increase the emulsion damage and would remove information. The whole point of digitising is to prevent future handling of the original objects. The procedures were set to keep the original file of the damaged negative as a master and to digitally improve on a copy. The broken negatives were relatively easy to restore, but the success of the restoration on the reticulated negatives was more varied. Often broad lines of emulsion were missing, and it was not possible to replace the lost information. However, when there was damage in the sky-area of a landscape-motive, removing these black lines in the sky was easy and greatly enhanced the image. Our objective was never to restore the image to the level of a new one, but only to remove visible noise, to make the image easier to read. There is nothing wrong with old photographs looking their age.

### *Digital printing*

Chemical contact-printing from the old negatives will result in very sharp, but often extremely high-contrast paper prints. It takes a great deal of photographic skill to make high quality prints from old negatives on modern photographic paper. The reason for this is that old negatives were adapted to albumen printout paper, which had a contrast-curve that suited these negatives. When a hard contrast negative is printed on to a very low contrast paper, this is a perfect match. Modern chemically developed papers and modern negatives have an opposite character: hard contrast profile in the paper, fitting modern low-contrast negatives. Printing old negatives on modern paper therefore gives information-loss in the lightest and darkest areas. A scanned negative, however, gives a completely straight contrast curve, the result being that the digital print will yield all the information present in the old negatives and will be a surrogate print closer to the original albumen print. Large digital printers also present rationalization benefits for printing large format exhibition-prints, which on photo-paper was a very labor-consuming and demanding work.



Tromholt, S: *Niels Andersen Sara, 56 years old, Koutokaeino, 1883.* (... 56 years old, Kautokeino). Print on modern developing paper from glass negative.



Same image, digital print from glass negative.

## Project “Jubelknud”.

### *Database, scanning for publishing on the web*

The next natural steps in the digitisation process were importing the files into a picture-database and publish them with a web interface and functionality. At this time, in 1999, the Norwegian library system BIBSYS had no picture-module, and although a museum-picture-database was in use, it did not have web-publication functionality. A grant from the Bergen European Culture City 2000-celebration and the University in Bergen made it possible to develop a searchable database with web-functionality, presenting Knud Knudsen's images from Bergen and Western Norway as a gift from the city of Bergen to its inhabitants. The chosen solution was an SQL /Access database with a web-interface. The deadline for publishing this was early 2000, which meant that we had less than a year to write the specifications, develop the database and enter the relevant images into the database in time for the Culture City opening.

It was specified that the database should be well adapted to web publishing, as we wanted to make database-generated net-exhibitions. To be able

to achieve this we needed the files to be the right size, not too big for searching and copyright-reasons<sup>1</sup>, not too small for reading the image-information. We also wanted to incorporate information from our old card-catalogue, our old signature-system and the hierarchical keyword and geographic system.

For the web-version we wanted a design that was easy to read, with functions of topic browsing, further search, marking and selection, and the possibility of ordering photographs by e-mail. The web-interface should also have aesthetic qualities suitable for the presentation of photographic images.

### *Economic considerations – time and labor*

The database was originally created for the Knud Knudsen-archive but we needed a strategy to integrate the new database with the rest of the collections and our daily scanning of new images. We did one more en-bloc digitisation of a collection of 1500 negatives, but soon realized that the scanning and registration of reference information into the database was so labor consuming that it would be unwise to continue with projects of large-scale en bloc digitisation. We therefore returned to the earlier practice of limiting the digitisation to ad hoc work on particularly endangered negatives, work in connection with exhibition-projects, printing for the reading room and digital production for the public. In addition, we used smaller scanners to make image-files that would fit the web-format, limiting the large camera-scanner exclusively for more demanding tasks and larger files. The database-digitisation and the larger scan for preservation became parallel workflows. The first could in principle be done by any department-employee on an office scanner, while the photographer made the high-resolution work with the scanner back of the large-format camera.

### **Project “Between two Wars” and “Women in the Workforce between 1916 and 1960”**

These two projects stem from the wish to present the collections in a wider historical setting that would combine different media-types and special collections of the Library: manuscripts, newspapers and photographs, historical articles, bibliographies and links.

It would contain articles on local-, social-, cultural-, political-, research- and economic history and we had hoped that this collaboration could include other local libraries, archives and museum. A project of this scale would need external financial backing and could not be carried out as originally proposed. But it lives on in a different form. The strategy is to propose more limited projects with

---

<sup>1</sup> The use of our images in printed form is subject to a copyright fee. The images in the database are therefore presented in a file size that is too small for unauthorised use in printed form.



*From the 1928 National Exhibition in Bergen. Photo: Knud Knudsen & Co.*

content that can be fitted into the time-frame of 1918 to 1929 and thus let the “Between two Wars”-project (Ill. 4) grow according to the budget situation. One of these financed part-projects is the “Woman in the Workforce”-project; another is a project of inner-city children.



*From the Janus textile company, 1930's. Photo: Atelier KK.*



*Children at Bergen Harbour, 1920's. Photo: Schumann Olsen, O.*



## **Present situation and the future**

### **Digitise how: Technology; limitations and possibilities.**

Our equipment was regarded as high end when we bought it for DIGIKNUD in 1997, and our procedures and priorities of the time was a result of technical limitations and possibilities inherent in this technology. Scanning large negatives with a very high resolution was slow and took up much of the photographer's time. As long as the large files were stored externally on CD's, it never became a storage problem. Technological development has now made the scanning technology better and faster and CD's are no longer considered a safe, long-time archival digital carrier. Our digitisation strategy has therefore changed with the purchase of new equipment and the development of new archival standards.

The new Hasselblad with a digital back does the high resolution scan on a fraction of the time used by our old camera. Therefore, large-scale projects can again be considered, the priority still being endangered collections first, then particularly interesting collections for publishing and other special projects, and the digitisation of units for our users for publication research or private use. The standard for digitisation is now 300 dpi with 4000 to 7000 pixels for the longest side of the image, 16 bits and tiff-format. For exhibition purposes, even larger files can be made. Storage carrier is now changed from CD to files on University servers and network. At this point we have approximately 250 Gb stored files, with a calculated yearly growth of approximately 20 Gb. This is considerable, and smaller museums and independent institutions may not have access to this amount of storage space. In addition to the master files, backup files must be produced continuously. At present, the IT-department automatically produces tape backups of all the resources on the institutional network. The optimal security would be obtained by storing an additional backup on a server placed in a different building.

### **Digitize when? Waiting for lower cost and higher quality**

With the extremely fast technical development in scanner and photographic digital technology, software development and storage capabilities, it is harder than ever to know when to change the photo-studio and wet-darkroom into a digital work-place. There are advantages and costs to being a pioneer, while the waiting game can be rather nerve wrecking.

The general trend in the development of camera and photographic techniques has been for the cameras to get smaller and easier to handle and with more automation, and for the film material to develop a finer grain-resolution and sharper lenses. This trend has continued into the digital age. The photo-technology

is still developed for the mass market, where easy handling and convenience relative to cost is more important than quality. The proliferation of low cost and automatic scanning equipment may lead to the conclusion that preservation-work can be done without professional competence, and could tempt museum and archives to start doing digitisation projects too early, with the rewards of high flexibility and low cost. This mistake has been made before. In the 1970's, there was a drive throughout Norway towards collecting old photographs by small local museums and private organizations. Re-photography projects were started by enthusiastic amateurs with no knowledge of how to choose the right film and do high quality development. The result was several so-called photo-conservation collections of very poor quality. However, with photographic expertise, high-quality equipment, medium to large format film and competent development, film-based re-photographing still does long-lasting and high quality job in photo content-preservation. There is virtually no limit to enlargement and detail information, and as a medium it is not dependent on a specific media-reader.

With the present digital technology, the popularity of doing film-based preservation-work is waning. The popularity of the digital concept may have drastic consequences for the quality of the output and its durability over time. The challenge is to jump onto the fast moving train at the right time, i.e. make the right decisions both in choosing the project, the software, the scanner & camera/hardware. Most digital equipment will be obsolete in 5 years time. Buying quality equipment is expensive but will be a guarantee that quality and standard is valid for a longer time. Low-cost scanner equipment has improved over the years, but for conservation purpose there will never be a low-cost solution. And last but not least, knowledge of photography is more important than computer- and digital knowledge; digital photography is still about photographic seeing.

The economic cost of pioneering is high. Our scanner-back and first printer were very expensive, and it needed a skilled photographer to run it. Because it was so expensive, we kept it long after it was ripe for replacement. Lowering our standard for the new investment was not an option, and at this end of the technology scale, the prices were as high as before.

The advantage, we learned, in pioneering such a radical technological change is the buzz of doing something new and exciting, being at the front where things happen and finding our own quality standards and work-solutions. There is always a danger of choosing the wrong strategy, and mistakes may be costly. We are comfortable with our choices when it comes to the technology and the digitising standards. This work still stands the test of the time because we never compromised on quality and expense.

### *Making our own database or waiting for the columbi egg*

Our decision on developing our own database-software may be more questionable. It was mainly a question of timing. We were ready for a database with a web-configuration before most other museums and libraries. A few commercial systems were considered, but they were either too expensive or they did not fit our needs. At this time, an image-database for museums was under development on a national scale. However, we were not in a position to have any influence on the priorities and choices made in that project, and a museum-database might not be what we wanted. A museum-based database will have its strength in registration and description, while a library-catalogue will put emphasis on search and retrieval and availability to the public. Besides, the web-extension was at the time rather basic and did not have priority. The advantage of developing our own system was, of course, that it became custom-made to serve the needs of our workflow and the very high output of user-service. Later the image-base has been fitted with administrative- and loan-functionality that has meant a great rationalization of our daily customer-service. The process of planning for the database-functionality, writing specifications to our needs and seeing all this come to life in the resulting database and beautifully designed web-application, gave us great satisfaction and pride. And through the years it has served us very well.

The disadvantages in making custom-made applications are however several. Any database and web-application needs to grow and develop with time. This has not always worked. Application-development is largely person- and budget-dependent and we no longer have the in-house knowledge to change the database to our present needs. We are now in the process of re-examining our options for the future. We hope that we may find a solution to keep a modernized version of the database, and to change the net-application into a more modern interface and more up to date functionality so it may become more easily integrated into our library portal. A national ABM search functionality portal does not exist at present, but our future system should prepare for this possibility.

### **Digitise what and how much?**

In principle, our whole collection is open to the public. Today this means availability on the web through a searchable database: the content and number of images in the database should correspond to the content of the full archive. The amount of files published on the web per date (spring 2006) is 13000, which is less than 2% of the full collection, so the database gives us less credit than we are due. (Our in-library database has 40.000 files including mini-registrations.) Digitisation of unpublished source-material has a high priority both in the University strategy-

plan and that of the University library. The enormous backlog in the photo-archive necessitates a strict priority-plan. Criteria for this should range from age and level of physical damage to content-value. From a conservation point of view it would probably be both most cost-effective and safest in the long run to employ a specialist in photo-conservation to do the physical day-to-day health evaluations of in the collection and thus pick out the photographs in most dire need of treatment and passive cold storage. A large amount of our collections are in good condition, and the source value is varied. Properly stored in our climate-controlled rooms there is no reason why such pictures could not be left un-digitised for decades yet. The digitisation for preservation would then be done side-by-side with the digitisation needed for communication and publishing projects and the interest of the general public. Unfortunately, in our case the University does not have expertise of physical conservation of photographs. The need of our users is often given precedence before the preservation needs. For a library, this is probably typical. In my opinion, the only answer to the problem is increasing our resources of manpower.

### **Why Digitise?**

Analogue or digital preservation– or maybe both

In photography-preservation, the baryta paper prints from film-based procedures of image-preservation were expensive, cumbersome, but of high quality and long lasting. Digitising procedures present new possibilities in dealing with historical sources, and can in some cases perform tasks better than our conventional tools. Digital conservation is vastly more convenient, in some ways better, (printing and digital repair) and is steadily achieving higher quality. For digital prints both paper and ink quality now approach archive standard. However, if done to the highest standard, the digital workflow and equipment is not a cheaper alternative. It has the disadvantage of being tied to reading- and storage media that have a short time-span guarantee, and it therefore needs future and repeated re-playing into new media and new formats. Also, the decisions once made on resolution and file size will for the future determine the possibilities to enlarge for detail. The optimal solution in this situation may still be using both analogue and digital strategies in photo-preservation, opting for the belt-and-braces-tactic. For most institutions, this is a not a viable solution.

## **Conclusion: Popular presentation and democratic access to the sources**

The introduction of the digital technology in the Picture Collection has taken place over several years and in a manner relevant to our growing knowledge and understanding of the new technology. As in most groundbreaking technological changes, the new technology is initially used for solving old problems and rationalizing old procedures. This is the stage of digitisation for conservation and content-preservation, for us the DIGIKNUD project. The next stage is to discover and start using the technology for cataloguing and collection-management, developing databases and improve public access to the collections and the individual collection elements and items. (Project JUBELKNUD) This is the work that ensures access to the pictures as raw sources, and is vital to the interest of the research-community.

The main advantage of the digital technology in a global sense is without doubt the advantages in dissemination and publishing knowledge. Earlier, source material was locked in storage rooms with limited access, but the digital revolution has had the democratic effect of giving the general public access to unique source material.

Lifting the historical sources into the public arena is important. However, an even greater and more interesting challenge is to go beyond the database-presentation of the sources, by making web-exhibitions, portal-presentations and exhibitions by topic, like the “Between two Wars”-project. (Ill 4) In an information society, we need to present our resources within a context, which is exciting and will appeal to people of different ages and interests. The internet-medium is well adapted to presentation of historical knowledge. The librarians and archivists of unique documents, together with the academic staff of the University, have the competence necessary to put together this kind of information in a way that can be a travel of discovery into the past. The different elements of a library-collection will thus be given added value through context. This is a modern publishing procedure that is more flexible than traditional publishing; it is a challenge to the library-employees and a reaching-out to the general public. It is, however, labor- and time-consuming work. It presupposes a spirit of collaboration between institutions and departments, and it may go beyond the task of a research-librarian as we see it today. But from the standpoint of a picture-archivist, this is without doubt the most exciting future use of the digital media.

## **Bibliography**

Den Norske Ikonografiske kommisjon (1953). *Redd fotografiene*. (Save the photographs.) Oslo.

Norsk Kulturråds arbeidsgruppe for fotobevaring (1992). *Fotografi - fortid og framtid. Plan for vern av fotografier*. (Photography – past and future. Plan for preservation of photographs.) Oslo.

Oulie, H, 2005. *Digitaliseringer av fotosamlinger*. ABM skrift no. 1 (Digitising photo-collections.)

# E-books and their future in academic libraries

By Susanne Mikki and Elin Stangeland

In this chapter we will take a look at e-books in the context of the academic library, and discuss important issues regarding experiences with use of e-books at the University of Bergen (UiB) and how this affects their procurement. We will also look at how the open access movement influences the e-book market before we end the article with a discussion concerning future directions of e-books as an electronic resource.

## A brief history of e-books

The first major efforts in the area of electronic books were started in the 1970s with Project Gutenberg<sup>1</sup> and the Oxford Text Archive<sup>2</sup>. Later in the 1980s and 1990s book vendors recognised the possibilities of providing content in digital form, typically published on CD-ROM or to be used with personal digital assistants (PDAs) (Tedd, 2005). At the beginning of the millennium special e-book reading devices such as the Rocket E-book reader were developed. These readers turned out to be lacking in user friendliness, and this combined with the fact that no standard format managed to break through, the use of e-book readers stagnated. These readers have made the terminology of e-books somewhat confusing, and many people have defined them by the reading appliance rather than the idea of a digital book (Lynch, 2001). An example of this is the definition made by Armstrong, Edwards et al:

Any piece of electronic text regardless of size or composition (a digital object), but excluding journal publications, made available electronically (or optically) for any device (handheld or desk-bound) that includes a screen (Armstrong et al., 2002).

For an academic library to facilitate distribution of e-books by using e-book readers is quite demanding, partly because of the issues in lending valuable reading devices but also in the problem of lacking format standardization. In the following text we have therefore chosen to use the definition created by the UiB e-book task force in 2003 (The University of Bergen Library, 2003).

---

<sup>1</sup> [www.gutenberg.org/](http://www.gutenberg.org/)

<sup>2</sup> [ota.ahds.ac.uk/](http://ota.ahds.ac.uk/)

“Electronic books” in this context is understood as books that may become available over the University networks (Internet) at ordinary computers (PD, Mac, UNIX) without special hardware or software (...).

### **E-books at the University of Bergen**

The e-book project was in 2003 given the mandate to investigate the e-book market, and evaluate the University of Bergen Library’s (UBL) future role in the process of acquiring, presenting and utilizing such resources at UiB (The University of Bergen Library, 2003). The backdrop of this mandate was the changing context for Norwegian academic libraries, which after the Quality reform of 2003 (Ministry of Education and Research, 2001) had to shift focus from supporting researchers to facilitating for students in learning centres. As a result of this project, UBL initially purchased access to netLibrary<sup>3</sup> e-books. Later ebrary<sup>4</sup>, Gale Virtual Reference Library<sup>5</sup>, Safari Tech Books Online<sup>6</sup>, Encyclopaedia Britannica<sup>7</sup> and others have been acquired. The experiences of these purchases will be discussed later in this chapter.

### **How are e-books used?**

In this section we will focus on our experiences with e-book use both at the University of Bergen and e-book use in general. As of yet e-books are not as established as e-journals in libraries, there are great variations both in access models and user friendliness of the offered services. Also, e-books tend to be more practical in some contexts, for example for literature research or looking up definitions etc. Still the digital media, mainly because of problems with user friendliness of reading screens, makes it challenging to read longer texts from cover to cover.

The everyday environment of students and researchers is to a great extent digital, and learning and teaching is simplified by making information electronically accessible. Those who are digitally literate are able to utilise the facilities of the digital media and have good chances to succeed with their work. E-books are part of this, assuming one has access to this media, which depends upon many factors:

- Electronic resources are, in general, expensive, and every subscription renewal or new purchase by the library has to be thoroughly discussed. These issues are revisited further in the purchasing section of this chapter.

---

<sup>3</sup> [www.netlibrary.com](http://www.netlibrary.com)

<sup>4</sup> [www.ebrary.com](http://www.ebrary.com)

<sup>5</sup> [www.gale.com/gvrl/](http://www.gale.com/gvrl/)

<sup>6</sup> [safari.oreilly.com/](http://safari.oreilly.com/)

<sup>7</sup> [www.britannica.com/](http://www.britannica.com/)



- Many vendors only sell access to a limited number of concurrent users, sometimes as few as one at a time. This means that users may be turned away from important resources, which can be frustrating when information is urgent, for example for students who frequently are given assignments with short deadlines.
- Vendors generally control access either by checking that the user belongs to a valid university IP-address. To gain access outside of university campus one has to access the resource by using a proxy server. This can be technically challenging and therefore limits the flexibility of accessing the services outside of campus.
- In the case of Ebrary, readers must install separate reading software. This is free, but requires an extra effort which could be avoided if for example a standard reading software had been used. On campus, the software comes as part of the software suite the university IT-department maintains, but for use outside campus you are obliged to install it yourself.
- For some users the various user interfaces of the e-book services can be confusing and therefore be restraining for their use.

For the remainder of this chapter, issues regarding how to make users use e-books, and their use are investigated: library teaching and marketing, reading abilities and habits, statistics and areas of success.

### **Teaching and marketing**

One important task at the library is teaching digital literacy (see Torras and Skagen this volume). Through making the electronic resources known their use will increase. At UBL we inform our users and train them:

- Courses in information literacy are held by librarians for students at the different faculties. This is a task of growing importance because they fulfil the student's needs in managing to navigate an increasingly complex world of information.
- Online and offline exhibitions on e-books are made, such as the Science Faculty's "Finn våre ebøker"<sup>8</sup> ("Find our e-books").
- When launching newly purchased products, information about this is published at the University news services, such as the library web-site, the student portal and the University online newspaper.
- Information about e-books is disseminated by giving out leaflets; which are put in the printed books when lent out.

---

<sup>8</sup> [www.ub.uib.no/avdeling/matnat/utstilling/eboker/eboker.htm](http://www.ub.uib.no/avdeling/matnat/utstilling/eboker/eboker.htm)

- The e-books are registered in the library catalogue, Bibsys<sup>9</sup>, in which users have the option to limit their search to electronic resources.
- E-books are also made available through the library portal. Currently this has special features for e-journals, but similar functionality for e-books is in development (Bakka this volume).
- The acquisition department arranges seminars on information of databases, where the target group is library staff members.

### **Reading abilities and habits**

The mean time spent in e-book reading sessions is generally between ten to twenty minutes (Rosy, 2002) indicating that users skim through the content rather than read in depth on screen (Woodward and Edwards, 2001). Inadequate screen readability may be one explanation of this; further e-books are often based on their printed counterpart and are therefore not always suited for on screen reading. Hopefully this will change as e-versions of books are adapted to fit the digital media.

Large parts of the UBL collections consist of electronic documents; particularly electronic journals are popular and in many cases available exclusively online. Still it happens that users ask for printed versions both for articles and books. There are several reasons for this: for e-books vendors often limit the flexibility of the digital format by restricting the user's copying and printing options. Access and using the resources are made cumbersome by supplying each page of a book as a separate digital document compelling the user to "turn" the pages rather than for example scrolling. Also, if the digital version is made from a printed original, the document quality can decrease. For example the quality of figures and pictures of digitised documents are in many cases low compared to the original, or they are separated from the original, forcing the users to access these on separate pages, which is inconvenient. Because of this printed copies of popular e-books is purchased, thereby limiting the versatility of the library collections.

Throughout the years there have been various efforts on producing hand held devices for reading e-books. So far these have not been successful for a broad market; technology is improving and more user-friendly products will be developed, but as for today cannot replace the printed material. From the environmental point of view it would have been a step forward if we could tackle digital information more exclusively without making printed copies for thorough study. Until then students and researchers prefer to acquire their own printed copies for note-taking, underlining and reading, and being accessible at any time (Mercieca, 2004).

---

<sup>9</sup> ask.bibsys.no

## **Statistics**

E-book user statistics at the UBL are, until now, inadequate. The data material is sparse, considering short time series and types of statistics offered by the vendors. Unfortunately, vendors measure different parameters such as number and length of sessions, queries, documents viewed, pages viewed, copied, or printed etc., which makes it difficult to compare services and usage. A potential strategy could be to perform surveys in collaboration with vendors; these combined with usability tests and focus group studies could serve to develop strategies for strengthening the library's collections.

Still, according to the statistic findings, there is no doubt that the use of e-books at UBL has increased continuously and is still increasing due to actions as listed above, and due to more frequent writing of assignments, where students are advised to use the library services. Increased usage of e-books is in accordance with international surveys. A five year survey carried out by Bailey, T. P (2006) shows a steady increase across several subject areas at Auburn University Montgomery Library. The usage of the printed collection, on the other hand, has declined during the same period, indicating that the electronic format is increasing in acceptance and popularity. De Rosa et al. (2005) confirm that college students show a high level of awareness of library electronic resources across all geographic regions examined.

## **Areas where e-books are a success**

There are some fields where e-books seem to excel, partly because they fit the activity at hand, or because their content is easier to access in a digital context. Important areas that we will discuss in this section are literature research, textbooks, reference works and grey literature.

It is common to use the library catalogue and databases as tools to retrieve information in situations where a student or researcher need a complete overview of a field. Many of these tools are difficult to use for non-librarians, and the thresholds of using them have made many users find alternative sources of information, or to manage with incomplete source material. Some of these problems have been solved with the access of full text documents because the users can now search the main source directly. The fact that full text is available on screen may save the researcher from reading long passages to evaluate quality and relevance of the source, but this depends on the amount of hits and the ability of the search technology to rate the results in a relevant manner.

Until now, only few textbooks used by faculty staff at the University of Bergen are available electronically. There are several reasons for this, the main one being that a large part of literature used at the University is not available in

electronic form yet. Particularly for subjects such as arts, humanities, law and social science many of the sources are written in Norwegian; and so far Norwegian publishers have chosen not to publish their books in an electronic format. For literature in science or medicine the literature is generally in English, however these textbooks are only rarely represented in electronic form and therefore few of these are available in the UBL collections. A possible explanation for this is that it is difficult for the librarians to get an overview of which e-book titles the various vendors offer. If a textbook is available online, it is also very often only available as part of e-book packages which are both expensive and where the majority of titles are of little interest.

Reference works are in many cases difficult to use, particularly in print form. This is due to the nature of these publications which often consist of alphabetically ordered overviews of topics or information organised by certain rules. The user has to figure out how to access the information before she can look it up. To search an electronic version of a reference work is in many cases more effective and this therefore makes them popular. By accessing the online version the user will also receive updated information since the publisher can easily revise the information with new material if relevant.

To get a more complete picture of the total amount of electronic literature available at UBL, also online material such as reports, manuals and journals should be mentioned. E-journals have long been established and are in many cases replacing the printed version. Students and researchers are becoming used to downloading articles at their own computers and are starting to show reluctance to visit the library to copy a printed article if the journal is not available online. In opposition to the practise for e-books, e-journal publishers set no limits for printing articles, and therefore make it easy to utilize the text as one prefers. The use of grey material such as reports, manuals and theses has recently increased. The main reason for this is the increased accessibility that the digital media allows by making a document available on the internet. Traditionally finding documents of this type is difficult even for trained librarians, but by depositing these on the internet, and making them available in search engines such as Google<sup>10</sup> or Yahoo<sup>11</sup>, the availability and hence the usability of these increases dramatically.

### **Extra material and services**

When discussing e-books, one issue that cannot be omitted is the extra functionality included for the users when accessing e-books. This can be various services or functionality such as the opportunity to:

---

<sup>10</sup> [www.google.com](http://www.google.com)

<sup>11</sup> [www.yahoo.com](http://www.yahoo.com)

- Look up references listed in a document
- Look up a word unfamiliar to you in a dictionary
- Translate selected phrases, if you do not understand the language used in the text
- Search an encyclopaedia for more information
- Access related maps or other multimedia materials containing background information
- Make notes or mark relevant text passages
- Add bookmarks
- Find related bibliographies and addresses to evaluate the author
- Cross search the library catalogue to find relevant additional sources
- Search the web to find more information about the subject
- Export references and citations to reference handling tools

These functionalities are included to give users some of the same options as they have when reading a printed book. Some of these are useful, involve the reader and make the research process more effective. The vendors seem to forget, though, that for users who face many different web-sites with information each day, storing notes etc. at various vendor web-sites is difficult to maintain and therefore impractical.

### **Issues concerning e-book procurement**

In this section we will discuss issues concerning the procurement of e-books based on experiences from UBL and also consortia collaboration on a national level. We will also look into free e-book alternatives such as are made available in institutional repositories, online reference services and other open access services currently available.

When the first e-book resources were purchased in 2003 the following factors were decisive for the purchase (The University of Bergen Library, 2003):

- Content and updates
- Price and opportunities for consortia collaboration
- User interfaces and search options
- Adaptability to library catalogue, import functions etc.
- Available statistics
- Technical requirements

It was decided to focus purchases on reference works and textbooks because these document types are as described above more user friendly in digital format than for example fiction literature (Woodward and Edwards, 2001). When it comes to pricing, the situation for e-books is similar to the situation for e-

journals, and buying access either perpetually or only for a period is expensive. As described above, e-book publishers are very restrictive when it comes to use, something which is unsatisfactory considering what they force libraries to pay for the content. Some e-book services require the user to install reader software on the users' computers and others require the user to login with username and password. Both of these involve overhead when it comes to administration and should therefore be avoided if possible.

After our initial purchase of 100 netLibrary titles and the finalizing of the Project Learning Centre 2003 (more about this in Tønning, this volume), e-books were treated as any other electronic resource. We continued to purchase single titles from netLibrary on a small scale, but it turned out that the administrative overhead of this combined with netLibrary's adding fees per purchase made it unpractical and expensive to keep up. Because of this the purchasing of e-books at UBL has mainly consisted of purchasing packages in collaboration with other institutions in consortia.

### **Consortia strategies in Norway**

According to Bostick (2001) several factors are important when libraries decide to collaborate in consortia. The most important is the possibility of sharing resources, both when it comes to access to the purchased literature, but also for administration and training. By allowing consortia staff to specialise on training and administrating the agreements, the local libraries are free to use their often limited resources to other suitable investments. By combining the buying power of several institutions, the consortia will be stronger when negotiating purchases from vendors.

Consortia work in Norway started in 1995 with the National office for research documentation, academic and professional libraries as the main facilitator. In the beginning only academic or research libraries were allowed to participate and the main goal was to cut costs. In 2003 the National office for research documentation, academic and professional libraries merged with the Norwegian directorate for public and school libraries together with other archiving and museum organizations into the Norwegian Archive, Library and Museum Authority (ABM). During this period the focus of what is necessary to purchase has changed from buying access to database services to purchasing full text resources such as e-journals and also more recently e-books (Sundby and Karlsen, 2005). In addition to the ABM consortium, Norwegian libraries also participate in other consortia such as the National Health library and PrioInfos Nord-i-KON offer.

Currently ABM has several e-book agreements; the most important of these

being on Safari Tech Books Online and netLibrary. In addition ABM facilitates several packages of reference works. Particularly the netLibrary consortium has been successful since the members by sharing their own books get access to everybody else's. Unfortunately the conditions for this consortium have changed, with publishers optioning to change sharing terms for consortia. As of January 1<sup>st</sup> 2006 netLibrary set a prerequisite for consortia that the libraries have to buy multiple copies of any document to be shared (Cook, 2005). The background for this is publishers concerns regarding the sales model where all libraries in a consortium are able to share the purchase of a single title. The amount of copies each member has to purchase depends upon the size of the consortium in question. The Norwegian netLibrary consortium for example consists of four institutions, and will therefore have to buy two copies of each title that should be included in the shared collection. For other larger netLibrary consortia this number can be six titles or higher and if we look at the average e-book cost which in addition to the initial print costs also includes a 50 % e-fee we realise that the costs of shared e-book collections are changing rapidly. As a reaction to this, the Norwegian netLibrary consortium has decided to freeze purchasing in the consortium until further notice.

ABM also facilitated the purchase of Safari Tech Books Online. Safari is sold on very different terms than netLibrary with each participant buying "points" which then can be traded into access to titles. The biggest advantage of purchasing access to Safari e-books as a consortium was the possibility of keeping down the amount of points that each library had to purchase initially. This gave each library the opportunity to test how such an access model would work, and evaluate use etc. without spending more money than necessary. Locally the advantage of Safari is that the books can be exchanged whenever necessary, so that the library is able to maintain its e-collections to be current and of interest to the library users. However the Safari exchange option makes it more challenging to administrate, particularly in regard to making it available in the library OPAC.

Finally the Nord-i-KON Ebrary consortium ought to be mentioned. Ebrary is sold in large packages with multiple user access and is currently only available in Scandinavia through the Swedish agent PrioInfo. This consortium, although very similar to a buying club<sup>12</sup>, currently has over 70 members throughout Scandinavia and the Baltic states.

---

<sup>12</sup> A buying club is here defined as a group of libraries purchasing resources from the same vendor like a consortia, but without the consortia benefits of common negotiations and administration, and potential shared access to the resources.

## **Pricing and licensing**

According to Bosch (Bosch, 2005) the pricing of scholarly communication through e-journal access is generally based on what the market will bear rather than the actual value of the commodity. One of the reasons for this is the publishers having monopoly on selling access to important titles and therefore chose to charge whatever they wish knowing that their customers are dependent of having access to the resource. The e-book market is less clear for purchasers because there are a big number of actors selling access, some being publishers others aggregators. Because of this the same titles can be available through different packages, or sold as single titles. For single titles the different vendors pricing practice varies a lot, with some vendors selling the e-version of a book cheaper than the print, while others add extra e-fees as high as 50 % of the print price.

A direct comparison between print and electronic versions is not completely viable though, because a printed book is inaccessible when in use, while an e-book is generally used for a shorter period, or if the license allows it, can be accessed by multiple users concurrently (Cox, 2004). Also the properties of the digital media allows for more flexibility in that the available full text is easier to navigate through searching possibilities, and also many vendors add extra functionality such as concurrent dictionary access. To navigate such a market is challenging because it is difficult to get an overview of available titles from different vendors and then to calculate the correct price.

Besides e-book vendors operating with a number of pricing options, they also have a number of different ways of organising the licence. The most common being site access for multiple users (Ebrary) or access to a fixed number of titles for a fixed number of concurrent users (netLibrary, Safari). The vendors also vary as to whether they sell the digital document (perpetual access) or whether they just lease you a right to access their online material. For perpetual access collections the library also has to deal with what happens if the vendor's web-site is shut down, or the library chose not to use this any more, in which case the library may have a digital copy containing the e-book text but no user friendly, long term storage solution to make it available to its users.

## **Future development**

In this section we will discuss various topics that will influence the e-book market in the coming years. An important trend is the development of free e-content. The open access movement for example organises free access to types of material which in many ways are competing with e-books, such as the Gutenberg



project<sup>13</sup>, Wikipedia<sup>14</sup> and other wikis, and the development of open repositories. Simultaneously Google, Amazon<sup>15</sup> and others are experimenting with the digitisation of books and possible new ways to utilise these.

As described in Jones (this volume) the so-called “Journals Crisis” has created large budgetary problems for the libraries. As a reaction to this, the open access movement evolved starting with the Budapest Open Access Initiative<sup>16</sup> in 2001 (Bailey, C. W., 2006) which defined open access as:

[...]By “open access” to this literature, we mean its free availability on the public internet, permitting any users to read, download, copy, distribute, print, search, or link to the full texts of these articles, crawl them for indexing, pass them as data to software, or use them for any other lawful purpose, without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself [...].

An important part of the open access movement is the development of open repositories, for example institutional repositories; in the Bergen case the institutional repository is the Bergen Open Research Archive (BORA)<sup>17</sup>. Here the institution itself takes charge of publications produced locally and by making these available online contributes to again attain the primary goal of scholarly communication which is to disseminate research results. As time goes by, and the institutional repository is accepted as the main archive for publications at the university, we see that many types of material which in many ways are parallel to e-books finds its way into the archive. The primary example is masters theses and PhD theses, but also other document types such as working papers, books (for example this one), are to be found in institutional repositories.

Another important development within the open access movement is Wikipedia and other types of wikis<sup>18</sup>. Wikipedia is a free online encyclopaedia that bases its existence on users creating the content. Any interested user can create new entries for a topic or edit existing texts. Changes and updates are thoroughly recorded, leaving each text with a traceable history which allows users to track changes and look at earlier versions of a text. Because of this critics claim that the content cannot be trusted in the same way as for example Encyclopaedia Britannica. This was disputed in 2005 by an investigation carried out by Nature where it was found that Wikipedia and Encyclopaedia Britannica contain about

---

<sup>13</sup> [www.gutenberg.org/](http://www.gutenberg.org/)

<sup>14</sup> [www.wikipedia.org/](http://www.wikipedia.org/)

<sup>15</sup> [www.amazon.com](http://www.amazon.com)

<sup>16</sup> [www.soros.org/openaccess/](http://www.soros.org/openaccess/)

<sup>17</sup> [bora.uib.no](http://bora.uib.no)

<sup>18</sup> [en.wikipedia.org](http://en.wikipedia.org)

the same amount inaccuracies (Giles, 2005). Wikibooks<sup>19</sup> is a sister project to Wikipedia where authors can write textbooks either together or by themselves.

Simultaneous with the open access movement another trend is emerging involving large companies such as Google, Amazon and even Microsoft creating and disseminating free digital content. Google, which mainly is known for its search solutions, is the main player in a new venture in collaboration with important American research libraries and also commercial publishers (Quint, 2004). The goal of the project is to digitise books both in copyright and in the public domain with the intention of making them available for both searching and in some cases in full text. Another similar initiative that were started right after Google as a collaboration project between Yahoo! , the Open Content Alliance<sup>20</sup> and later also MSN aiming to digitising public domain books (Dye, 2006). The planned models for giving access varies from only allowing users to see text snippets, to allow the user to access the whole or parts of texts either for free or for a fee. In the context of an academic library plans on allowing access to books that are currently out of print is appealing. However many copyright issues arises, for example Google's strategy of forcing publishers and authors to actively inform Google that they want to reserve their books from the planned scanning are questioned since this requires the copyright holders to actively protect their legal rights (Band, 2006).

Another interesting organization planning to allow access to e-books online is Amazon. Amazon has been digitising books since 2001 for its "Look inside the book" and later "Search inside the book" programs and therefore already holds large amounts of full text in storage. Amazon is also trying out publishing with their service Amazon shorts<sup>21</sup>, where they publish short literary works for \$ 0.49 each. This development will be extended with various sales models for selling access to e-books and Amazon are currently in the process of making arrangements with publishers. Amazon Pages are planned to allow full purchase of online access to all or portions of books, and Amazon Upgrade will allow customers who buy a printed book to add online access for a set fee. Currently Amazon are not planning to facilitate institutional access though, so it may yet be some time where for example academic libraries can add to their collections from Amazon (Quint, 2005).

As for UBLs regular book vendors, these have also started to look into the e-book market. The vendor Dawson books<sup>22</sup> for example, offer books in a number of different ways. In addition to selling print books, they are collaborating

---

<sup>19</sup> en.wikibooks.org

<sup>20</sup> www.opencontentalliance.org/

<sup>21</sup> www.amazon.com/exec/obidos/tg/browse/-/13993911/104-6954373-1206305

<sup>22</sup> www.dawsonbooks.co.uk/

with eBook Corporation<sup>23</sup> offering eBook Library (EBL), a platform for lending or buying e-books depending on the institutions need. Other vendors such as Delbanco<sup>24</sup> and Starkmann Library Services<sup>25</sup> are also in the process of establishing services for selling e-books.

In his article from 2001 Lynch states:

“Digital books are as yet only dimly defined, and will be a continued focus for the creativity and ingenuity of present and future generations of authors, teachers and scholars” (Lynch, 2001).

Five years later we believe that this is still the case, but would like to add that the publishers, at least for now, seem to influence the way that e-books are developing. Currently the publishers are caught between their wish to create good products for their customers and not loosing any of their perceived potential of revenue. Because of this most e-book services currently available are either impractical both to use and administrate, or they are so expensive that a potential purchase needs to be carefully considered. For some reason book publishers have very different views on digital dissemination of publications than journal publishers have, which allows users flexibility both when it comes to access models, and printing and copying options. The background for this could be the book publishers seeing an opportunity of selling both versions to their customers with increased revenue. The e-journal market has moved beyond this discussion as the institutions are starting to opt for e-only strategies. Such decisions are partly made based on the characteristics of the journals, mainly these are short to read and simple to print, and storing printed versions of these are not particularly cost-efficient seen from a library point of view.

Whether these arguments should be used for e-books is still open for discussion, the technological development may also affect this. As of yet we operate in a market where there are no reading devices available that are good enough to replace the user friendliness of paper, but according to Dick Brass, formerly a Microsoft executive, a book reading device of equal popularity to the iPod will in the future replace paper as we know it as the main medium for books (McCrum, 2006). If, or when, this happens the whole use pattern of books at least in an academic context will change, and maybe for the first time we'll see the paper free library. This can only happen when one issue is dealt with though, and that is the question of preservation. As mentioned in Jones (this volume) many institutions around the world are currently working with preservation issues,

---

<sup>23</sup> [www.eblib.com/](http://www.eblib.com/)

<sup>24</sup> [www.delbanco.de/](http://www.delbanco.de/)

<sup>25</sup> [www.starkmann.com/](http://www.starkmann.com/)

many of which remain unsolved however, and therefore the printed material is probably safe a while longer.

## **Conclusions**

In this chapter we have described how users find e-books challenging to read because of inconvenient user interfaces, use restrictions and uncomfortable reading devices. Simultaneously libraries are challenged by the fact that e-books are expensive, complex to administrate and with access rules and sales models continually changing. Also many of the use models that are offered today should be avoided, because they undermine the possibilities the digital media allows, for example to let several users access a resource at once. Nevertheless e-books are becoming increasingly popular and institutions such as UBL continue to purchase them. An explanation for this can be that even though the services are inconvenient, the advantages of being able to access these digitally outweigh the problems. In other words e-books are in the process of becoming an important information asset for libraries. It is therefore important that libraries around the world take charge and collaborate in developing strategies for e-book purchasing in the future.

Some of these issues may solve themselves however, as the open access movement and the wiki-concepts become more influential and actors such as Google and Amazon enter the field. Traditionally Google and Amazon are known to have policies that are less restrictive when it comes to what their users are allowed to do, and if they chose to apply these also for e-books, the consequence could be an increase in use flexibility for e-books in general.

## References

- Armstrong, C., Edwards, L. & Lonsdale, R. (2002) Virtually there? E-books in the UK academic libraries. *Program: Electronic library and information systems*, 36(4), 216-227.
- Bailey, C. W. (2006) What is Open Access?
- Bailey, T. P. (2006) Electronic Book Usage at a Master's Level I University: A Longitudinal Study. *The Journal of Academic Librarianship*, 32(1), 52-59.
- Band, J. (2006) The Google Library Project: The Copyright Debate. *OITP Technology Policy Brief*.
- Bosch, S. (2005) Buy, build, or lease: Managing Serials for Scholarly Communication. *Serials review*, 31(2), 107-115.
- Bostick, S. L. (2001) The history and development of academic library consortia in the United States: an overview. *The Journal of Academic Librarianship* 27(2), 128-30.
- Cook, V. (2005) NetLibrary Update.
- Cox, J. (2004) E-books: Challenges and Opportunities. *D-Lib Magazine*, 10(10).
- De Rosa, C., Cantrell, J., Cellentani, D., Hawk, J., Jenkins, L. & Wilson, A. (2005) Perceptions of libraries and information resources. A Report to the OCLC Membership. Dublin, Ohio, USA.
- Dye, J. (2006) Scanning the stacs: The digital rights issues behind book digitization projects. *EContent*, 29(1), 32.
- Giles, J. (2005) Internet encyclopaedias go head to head. *Nature*, 438(7070).
- Lynch, C. (2001) The battle to define the future of the book in the digital world. *First monday*, 6(6).
- McCrum, R. (2006) E-read all about it. *Guardian Unlimited*.
- Mercieca, P. (2004) E-book acceptance : what will make users read on screen?
- Ministry of Education and Research (2001) St. melding nr. 27. (2000-2001). Gjør din plikt – Krev din rett. Kvalitetsreform av høyere utdanning.
- Quint, B. (2004) Google and Research Libraries Launch Massive Digitization Project. *Information today*.

- Quint, B. (2005) Books Online: The Fee versus Free Battle Begins. *Information Today*. Medford.
- Rosy, R. L. (2002) eBooks for Libraries and Patrons: Two Years of Experience. *Liber Quarterly*, 12(2/3), 228-233.
- Sundby, O. & Karlsen, S. I. (2005) Fellesskap uten sosialisme : konsortiearbeidet i ABM-utvikling *Bok og bibliotek*, 72(4), 42-47.
- Tedd, L. A. (2005) E-books in academic libraries: an international overview. *New Review of Academic Librarianship*, 11(1), 57 - 79.
- The University of Bergen Library (2003) *Ebøker ved UIB*, Bergen.
- Woodward, H. & Edwards, L. (2001) E-Books (eBooks) Working Group: Shaping a strategy for e-books: an issues paper.

# Implementing the Digital Library – some theories and experiences on leadership of change

by Ane Landøy

## Introduction

Implementing the Digital Library means a lot of changes in the library. Some may be greater than others, but however we look at it, for many of our patrons the library will change beyond recognition, as will it change for the library staff.

In this chapter I will present some theories that I find relevant and helpful in change processes. Interspaced with the theories I will give a few examples of our experiences in the University of Bergen Library. More about the processes may be found in other chapters in this book. I then take a brief look at the leader's role in the different processes, and use implementation the Digital Library as an example. The implementation of the Digital Library will be seen as a change that may be planned and anticipated, as opposed to driven by uncontrollable "forces".

There are many definitions of "the Digital Library", but for our purpose it will be sufficient to maintain that it means a major change, both for patrons and staff, that it involves technological changes, and that it will concern most of the processes in the library.

## Theories on organizational changes

"Explaining how and why organizations change has been a central and enduring quest of scholars in management and many other disciplines." These are the opening words in an authoritative and widely quoted article from 1995, where Andrew Van de Ven and Marshall Scott Poole conducted a literary review of 200 of the most central articles about organizational change and development. In doing this, they introduced four basic types of process theories that explain how and why change unfolds in social entities: life-cycle, teleological, dialectical and evolutionary theories.

"Life-cycle theories" suggest that change has to happen in a logical and programmatic way, in which organizations will be moved from one point towards a prefigured end. Typically, change events in a life-cycle model follow a specific sequence of stages which are cumulative and related. Each change sets the stage for the next (Van de Ven and Poole 1995:515).

The teleological change theories see organizational change as proceeding towards a goal. The organization is seen as purposeful and adaptive. It can formulate goals, take action to reach them and monitor the progress. Development is a repetitive sequence, but not preordained (Van de Ven and Poole 1995:516).

The dialectical process theories explain change as a struggle of balance between opposing entities within or without the organization. We can speak of a thesis – antithesis – synthesis process (Van de Ven and Poole 1995:517).

The evolution models explain change as a “recurrent, cumulative and probabilistic progression of variation, selection and retention of organizational entities”. (Van de Ven and Poole 1995:518)

## **Libraries as organizations**

Academic libraries, or libraries that are part of academic institutions, have a number of characteristics that set them apart from other organizations; sometimes also from public libraries. Some of the characteristics of Norwegian academic libraries/university libraries that may influence the way they change may be:

- The library is organizationally placed within a University organization. This means that changes within the University may affect the library.
- Forces of change from the external environment, the society, that affect the University, may also affect the library.
- The library belongs to an academic institution, which means that the main focus is on furthering learning and research. This differs from the much broader aim of the public library, which also caters to the reading public for whom reading may be a pleasure in itself.
- We may also expect the library to see itself as taking part in the research process, meaning that the level of help and service given to an individual researcher will be high. This very high level of service orientation is of course a common feature among all librarians.
- An academic library of a certain size will be a hierarchally structured organization.
- It will be an organization with detailed and established routines for dealing with its patrons.
- The academic library is good at documenting, by statistics and written plans, prognoses and reports, what is going to happen or has happened in the organization over the time.



The German sociologist Max Weber developed a theory on leadership in different types of organizations, and found that the bureaucracy has its basis in laws, and is characterized by a specializing in responsibility, with competence and specialization as one of the fundamental criteria for the personnel, a clear and hierarchy of powers with many stages, and by written, formal, standard and repetitious task formulation. We may, from the above, see the academic library as a bureaucratic organization in Weber's sense of the word. (Weber 1979)

## **Change in organizations**

A definition of change in an organization can be that change will be registered when we see a major difference between two points of time, but there has been a relative stability in the situation at the two points. (Jacobsen, 2004) We can also consider that implementing the Digital Library will ideally be a change that may be planned, and therefore also a process that may be led by someone, either the formally appointed leader, someone delegated the task, or someone who will take the leadership and assume the responsibility.

## **The powers of change**

What are the driving forces behind change?

One such driving force may be summed up in "a feeling of imbalance" e.g. between two or more internal elements in the organization. This may be a negative imbalance – in which the organization has to correct a deficiency – or a positive imbalance – when the organization aspires to become even better.

Another driving force may be a measurable imbalance, concerning changes among the customers, the competitors, the suppliers or change in political issues etc.

In the library, we may see imbalances stemming from technological changes, where modern technology is brought to the market and the library has to use it. We may see imbalances when it comes to economic changes, when the subscription prices of journals rise a lot higher than the mean price indices.

The different driving forces may cause changing behaviour in the organization, but do not necessarily do so, and the same forces may have different impact in similar-looking organizations, from a total make-over to no action at all. For change to take place the driving forces must be translated into action, and must be considered relevant for the organization. Organizations will perceive the changes differently. Also they choose different strategies to meet the changes.

## Case Study: University of Bergen Library

### **Digital library = modern technology**

In an analogue library, patrons will find and retrieve material of interest, and the library itself will not be involved in their use of the material. In the digital library, patrons will retrieve and manipulate content as far as the functions and services of the digital library allow. At the same time we see that building a sustainable digital library presumes that the systems and services meet the need of some user community (Borgman 2002), in our case the scholars and students at the University of Bergen.

The cluster of changes that may be called “Implementing the Digital Library” consists for the most part of technological changes, and the cultural changes that come with modern technology and new needs rising from modern technology. One example may be Internet and its possibility of information overflow that leads to a new need for teaching of information literacy. This is a new need for students, but also for the library in order to stay current. See Tønning this volume, for more about this.

Among the technological changes in the environment we see the same ones that have led to many of us using internet banking, or reading the newspapers on the internet, that is, the spreading knowledge and use of personal computers, and of the web or internet. These changes have been fully documented, and here will be taken as a given.

Within the University the technological changes show themselves in the growing number of different Learning Management Systems being used in university teaching and learning over a period of time.

In the University of Bergen Library, the technological changes manifested themselves first in the building of an electronic catalogue, and purchase of computers for patrons to use for searches. Then came the purchase of electronic resources in order to make them available to patrons. There are a rapidly growing number of journals, databases, reference works and other electronic resources on offer, both for current and back issues, in many different forms, shapes and formats, and at a widely differentiated price range. A new library portal for accessing the electronic resources has been acquired. Computers for reading journals in the library were introduced, and later they have been augmented and equipped even for students writing essays.

In recent years the discussion about digitising the library’s own material (pictures and special collections of old and rare books and manuscripts) has been taken up. Along with this came a wish stemming from the Open Access movement, to create our own Institutional Repository.

The most recent aspect of the Digital Library to be implemented is the

developing of a set of courses in information literacy, also available on the internet.

All these changes are explored in fuller detail in the other chapters of this book.

## **What needs to be changed?**

Evidence from recent user surveys in academic libraries shows that:

- People trust libraries and their information
- The faculty think that they will be more dependent on electronic information in the future
- Patrons want to navigate the world in a self reliant way.

(Cook, Colleen 2006)

When patrons' behaviour changes, among other things because they become more used to computers and the internet, the library also has to change. Generally speaking, the changes may include changes in what an academic library building may contain: We may have an electronic catalogue; freely available over the internet from the workstations of student or academic staff. We may have several types of PCs for student use; both for single, double or group use, or for classes. We may have a wireless network allowing patrons to work on their own lap-tops all over the library building, where ever there are relevant resources.

Other current changes include what is on offer in the library or from the library. This may include electronics journals, databases and reference literature. Also, we may find the library involving itself in teaching Information Literacy to students, in cooperation with the academic staff at the university.

These sets of changes will involuntary lead to another change: a change in the staffing of the library, where we may experience a growing need for more computer specialists and pedagogues than traditional librarians; or a growing need for traditional librarians with an added expertise in computers and pedagogy.

Among the things that will stay the same we find the library's dedication to service, to academic research and to good value for money. We also believe that the physical building with the printed books and some journals will remain.

## **How are we changing?**

### **Electronic catalogue**

One way that University of Bergen Library has changed, is by having an electronic catalogue of our holdings, freely available on the internet. (Åsmul,

this volume) Thus, patrons may access our catalogue from their workstations, from their homes, or from all over the world, at their leisure. They may compare our holdings to those in other libraries that also are available on the net. As a result of this we experience that our users have a growing demand for inter-library lending of books that are found in the catalogue. Of course, we will at the same time experience a reduced demand for inter-library lending of journals, when they are more easily accessed electronically.

### **Purchasing electronic resources – what about the paper?**

Another way our library changes is by purchasing electronic resources. Again, patrons may compare our holdings to those of other libraries, and decide if we have the relevant resources.

When we purchase electronic versions of resources and we already subscribe to the printed version, the discussion arises whether to keep the printed version in the library, or discard it. Some of our patrons may want us to keep the printed versions, but if we decide to keep them this will take up space in the library, and render it less flexible. If, on the other hand, we decide to throw the paper away, we will have more room in the library, for instance for workstations, where patrons may search and access our electronic holdings. This problem is being solved differently in the different libraries, depending also on space constraints.

The libraries with a certain amount of electronic resources are finding that the printed versions will not be in demand, and not be used. This seems to be the result of patrons starting to use electronic resources, and then falling out of the habit of using printed matter. We also see that research groups with sufficient funding will pay for single articles themselves, if the library does not subscribe to the actual journal. If the library, for economic reasons, decides not to subscribe to a certain electronic resource, it runs the risk of becoming obsolete for this particular group of researchers (Sivertssen 2006)

An other experience from the academic libraries that have many electronic resources is that the academic staff is happy to access the electronic resources from their office computers. Often, they cease coming to the library building at all. The library loses the day-to-day contact with one of the major groups of patrons, and thus a possibility for spontaneous feedback, be it positive or negative. The present spontaneous feedbacks will probably mainly be “I can’t access this resource, what’s wrong?” or “Why doesn’t this work?” (Sivertssen 2006, Kongshavn and Sivertssen this volume)

In addition to the changes we have seen, we experience a growing need for instruction, teaching and information, especially about the electronic resources, for many kinds of patrons. These needs will be diverse, and will need more or less tailor-made courses for each different group of patrons.

## **Teaching Information Literacy to students**

In Bergen we have experienced, and experimented with, the teaching available from the library. We have included Information Literacy among our services, and have, as already mentioned, different courses for different groups of students and academic staff. For the first term or first year of study, students will typically be offered a course in how to use the library catalogue, a few chosen electronic resources, about academic integrity and how to cite, and a guided tour in the library. For more advanced students we offer courses that focus on more advanced search possibilities among the electronic resources, and for masters, doctoral and academic staff we offer courses in electronic resources that are subject specific, and in reference management systems.

The different libraries have diverse student groups. To a certain degree they offer different electronic resources, depending on subject. This makes the course menu different in the various libraries.

Another difference is the degree of contact between the library and the departments when it comes to encouraging their students to follow courses at the library. Sometimes the contact is good, and the library courses appear in the student's schedule through the learning management system, while we have greater challenges in getting students from some other departments to come to the library courses.

For all the libraries we see that the best strategy is to establish and maintain close contact with the departments, both in order to make use of their help in promoting the library courses and also for the library to be able to give targeted courses that suit the academic work that the students are doing at that particular stage in their education. In this way we can show the departments the value in library training for their students.

In addition to the close contact with the departments, a strategy for teaching Information Literacy at the University of Bergen has been developed by the Library, and this strategy has been discussed at the top level of the University itself.

The teaching initiative is partly a result of implementing the digital library, but also because of changes in Norwegian Higher Education. (Tønning this volume), (Skagen and Torras this volume)

## **Digitising our own collections**

In addition to purchasing electronic resources in the commercial market, we have decided to digitise documents that are in our own holdings.

It remains to see what the most useful starting point is for this work. Would it be texts for study purposes, or original sources, like legal documents, for

research? Would funding be available to digitize a part of the holdings, perhaps a special collection that would benefit from digitization? In Bergen we are at the moment trying all these strategies in different projects, and have not yet evaluated this fully. See also Greve and Kyrkjebø, both this volume.

### **Library staff – keeping ahead of the development**

Library staff will need to work hard to keep up with the electronic world. We need to be aware of the different search functions, and how to make them interact with the interface we already have in the library. We need a good interaction with the academic staff, in order to keep them informed, and promote certain resources and courses. We need to keep an eye on the statistics, making sure that we get good value for money. Also, we need a certain level-headedness, in order to not be seduced by all the different offers in the market, but to be able to distinguish the useful from the useless resources.

We also need to keep abreast of the pedagogical challenges, with more and more courses being taught about more and more different subjects. It will be necessary to evaluate our use of the different kinds of personnel resources. When we decide to use new groups of personnel in the teaching, or to teach differently, there is a need to be trained.

### **Incremental changes or a revolution?**

When we look at the examples of changes, we see that the changes seem to be incremental, small and spread over time. Indeed they became more like adjustments than changes, once we started cataloguing electronically, and once we bought our first electronic journal.

At the same time, we can see that the library has changed beyond recognition over time. Imagine what the library was like before electronic catalogues were freely available on the net: patrons needed to come to the library building and use the catalogue themselves and searching in the catalogues was limited to author or subject. Ten years ago only printed journals were available and the researcher had to come to the library to read or copy from them; compare this to today, when the latest research arrives at the researcher's computer almost by itself.

The revolutionary aspect is the change in medium, from paper to electronic. Can we compare this to any of the other great changes in library history? From scrolls to codices, from hand-copied to printed books, these are changes considered by many to be a revolution in the book – and in library history, but in my opinion the embracing of the idea of the digital library must be considered even more of a revolution.

In this changing library, the library staff to a much larger degree remains constant. The same kind of people who used to type the catalogue cards now catalogue electronically. The same kind of people who used to show patrons how to search by author name in the catalogue will now teach them how to search all the possibilities in the catalogue, and also in all the electronic resources. (Åsmul this volume)

### **Case compared to theory**

Returning to the typology offered by Van de Ven and Poole (1995) we will see that the changes forming a part of the implementation of the digital library may be classified more easily under some theories than under others.

It will be difficult to see the digital library as a part of a life-cycle for libraries, where the digital library is like a seed, waiting for its time in the sequence. The change is too radical and too different for this to be considered a natural development, or following a law of nature.

The teleological theories look at planned changes, where the organization proceeds purposefully and adaptively towards a goal or an end state. It is easier to see the implementing of the digital library within this framework.

The dialectical process theories concern themselves with the balance of power between maintaining the status quo and opposing forces. We can easily imagine that this also has been the case when it comes to implementing the digital library.

Finally, we have the theories that Van de Ven and Poole called “evolution”, where change is seen as a recurrent, cumulative and probabilistic progression of variation, selection and retention of organizational entities. For some aspects of the digital library these may come into play.

Van de Ven and Poole (1995) then went on to classify the different models using the dimensions “units of change” and “mode of change”. If we follow this classification, we find that implementing the digital library, being a change that takes place in one organization and with a certain degree of prescription to the mode of change, fits among the teleological theories.

### **What kind of change?**

Changes that may be classified as teleological point towards a planned change (Jacobsen 2004:35). For the planned change there are certain kinds of leadership and leadership strategies that have proven to be more fruitful than others.

Generally speaking, we would first need to examine the limitations to the

decisions we are able to make. Often, we will start by making a SWOT-analysis, where we examine the Strengths, Weaknesses, Opportunities and Threats to a given organization, at a given time. The strength and weakness sections form an analysis of internal issues, while the opportunity/threat dimension comes from an external analysis in the relevant environment.

All action is a result of cooperation between the individual and the environment. We will see that it is often easier to change the environment than the individual.

Jacobsen (2004) describes two different strategies for leading planned changes, quoting from Michael Beer and Nitin Nohrias book “Breaking the Code of Change” (2000). Stating his fundamental assumption – that a planned change is possible if a) the correct strategy is chosen, and b) the process is correctly led and suitable to both the process and the situation (p.193f) – he then goes on to describe the two strategies. One he calls strategy E, or Economic, and the other is called strategy O, for Organizational.

Strategy E is used to create an added economic value. The focus is on formal structures and systems. It is run by the top leader with much help from external consultants and by using financial incentives. It is planned and programmatic (Beer and Nohria 2000 p.3, quoted in Jacobsen 2004).

Strategy O, on the other hand, is concerned with developing the human resources in the organization, making them capable of implementing strategy and learning from earlier experiences in processes of change. The focus is the development of a culture that will in turn create enthusiasm. There is a large amount of staff participation, and consultants and monetary incentives are less frequently used. Change happens incrementally, and less planned and programmatic (Beer and Nohria 2000 p.3, quoted in Jacobsen 2004).

The two strategies are different, particularly for the following questions: 1) What kind of goal? 2) What is the top leader’s role? 3) What is changing? 4) How is the change process planned? 5) How is motivation created? 6) What kind of role do external consultants have in the process? (Jacobsen 2004)

In the Digital Library-setting, we can see that strategy E may be the correct strategy to take when approaching the major decisions involved in implementing the Digital Library: the decision to start cataloguing electronically; the decision to convert the catalogue from print to electronic; the decision to purchase electronic resources systematically in order to create electronic holdings; the decision to create an institutional repository; the decision to digitise a larger amount of the library’s own holdings. The top leadership needs to guarantee this type of decision, which concerns major investments, both monetary and in terms of labour. Other decisions, such as “how do we do this?”, “what now?”, “how can we use this aspect of the Digital Library to better our service to customers?” may



well be better off with a strategy O process, where the library staff will themselves be much more pro-active. Especially when the goal of a process is to change the culture, not routines only, a strategy O process seems to be the most functional.

### **What kind of leadership?**

From the differences between the two strategies, we can easily see that they will require two different styles of leadership.

While strategy E focuses on economic goals, strategy O focuses on cultural changes that have to happen in the organization. For the library the economic goals will be securing funding, considering how to use resources (both monetary and human) efficiently, and what kind of modern technology to implement when. As for the cultural changes, they could be new ways of meeting customers, new products and services, or new routines and systems that have to be developed, and they are changes that should be learnt by the whole organization.

In strategy E processes the top leadership has a key role. They are the ones that can commit the organization and a certain amount of the resources to a certain course of action. They are the ones to formulate the problem, the goals and the solution. The changes that follow from this process will often concern themselves with systems and structures. The top leadership will still have to gain acceptance in the organization in order for the changes to be implemented, but this may be done by threats (of job losses or bankruptcy) or promises (of bonuses or gains). Often external consultants are heavily involved in at least some phases of the process. (Jacobsen, 2004)

In processes following strategy O the goal will often be to achieve an organization that is more capable of learning and developing. Systems learn by focusing on staff and groups, and humans may be lead by also being allowed self-leadership. Strategy O assumes that learning is a never-ending process, and that learning happens by correcting mistakes, and then reflecting on our behaviour in order for the same or similar mistakes not to happen again. The top leadership's role in this kind of processes is to create a vision and an enthusiasm for learning and development in the organization. They must delegate, participate and support. Everybody must be allowed to bring in ideas and viewpoint. In order to implement, most, if not all, employees have to agree to a course of action. (Jacobsen, 2004)

This type of process focuses on the humans in the change processes, including cultural aspects like competence, values, emotions, relations between people and groups of people, and the ability to cooperate and to handle conflicts. The informal elements will first change, and then they will influence the formal

elements of the organization. This is typically done in smaller instalments, as incremental changes, rather than in all encompassing sweeps. Motivation for change may be created either by focusing on a dissonance in the present situation, or by focusing on aspirations to become even better. The process of change itself must be motivating, in order to create energy. The leader of the process, who may or not be the top leader, will have to inform, involve, and inspire.

In the Bergen context we may see that while converting the card catalogue into an electronic catalogue, freely available on the internet, was a strategy E process, the fact that we now catalogue all electronic resources with the central UBB-signature instead of the signature of the library that pays for it, or where it “belongs” according to patron preferences, may be seen as a result of a strategy O process.

Also, we may view the original decision to purchase the Knudsen picture collection in the same light, where the purchase itself came as a result of a strategy E process, and the development of the digitising strategies has clear indicators of being a strategy O process.

Once a library has decided to look into implementing the Digital Library, the next step will be to identify the different levels of decisions as belonging to either an E- or an O process. Accordingly the responsibility for the processes will be placed either at the top leadership or the middle manager level. What would then be the best way for the leader to act?

At the start, when the decision to go in the digital direction must be taken, there may be a myriad of questions to answer and details to look into. Depending on the level of digitisation already attained in the library, and how far into the process the organisation has already stumbled by coincidences, the leader will need to address questions like funding, technology, what parts of the digital library to implement, and in what order. The leader needs to make sure she has all (or as many as possible) relevant facts, and will get them by employing external and internal experts, find descriptions of best practices from other that have already been through the same process, by assessing her own library’s strengths and weaknesses, and by communicating with relevant others, especially the ones that deliver the funding. In some stages of this process external consultants may have a role.

Once this process has passed a certain point, typically the task forces for different projects will be created, or the task of implementing will be handed over to the relevant part of the organization.

At the University of Bergen Library, we see this for instance when the University Library building was renovated and made into the Arts and Humanities Library building. Once the decision to rebuild had been made by the Library, and the funding was agreed upon by the University, a task force was

created with members of the most relevant groups, to plan and implement the building, moving and change process (Bagge this volume). This building process is also an excellent example of a process where staff was highly motivated for the great change, both because of the perceived drawbacks with the old library, and also because of the anticipated gains to come in the modern building, both for patrons and staff.

In this change process, there are also some limitations arising from the fact that the organization is classified as a bureaucracy. From what is written about this, we will expect change in a bureaucratic organization to be mainly top-down, mainly concerned with structures, systems and routines, and the entity will often be considered as quite difficult to change. But in an academic library one of the core values will be learning and development, and there are a lot of information specialists, who will eagerly follow development within the academic library sector. Also, the level of service orientation will probably be a lot higher than for an ordinary bureaucracy. Both these factors play key roles in determining the probability of the university library to manage a high degree of changes and learning.

In the task forces, wherever in the organization they may be located, however they are put together and whatever part of implementation they are focusing on, there are different kinds of smart behaviour from the task force leader, which will make the work to go as smoothly as possible.

In the process of setting goals leaders can influence the levels of self-efficacy in the individual members of the task force, by influencing their ideas on how they can perform, both when it comes to the levels of difficulties they can handle, and their general feelings of competence. These are factors that have been shown to influence performance by individuals. Leaders may also help members perform better by offering opportunities for training and by making relevant resources available. Thirdly, leaders may, by delegating responsibility for parts of the task, empower members. (Strand 2001:457)

Leaders also have a responsibility for integration and culture building. For the leaders to be moving the library culture from that of being an analogue library to a library that also is digital means hard work. In the University of Bergen Library, this means creating acceptance for this change in the face of a lot of different sub-goals, which different parts of the library organization or the University itself have. The main responsibility for this falls on the top leadership, but also on the middle management.

We need to prepare the organization for the fact that the different changes we have seen so far are only the beginning.

## References

- Beer, Michael and Nohria, Nitin (2006) *Breaking the Coche of Change*. Boston, Ma: Harvard University Press.
- Borgman, Christine (2002) Challenges in Building Digital Libraries for the 21<sup>st</sup> Century. *Lecture Notes in Computer Science*, 2555, p 1-13.
- Cook, Colleen (2006), "LibQUAL+™" at the 8. International Bielefeld Conference, Feb. 8<sup>th</sup> 2006, retrieved from [http://conference.ub.uni-bielefeld.de/2006/docs/presentations/cook\\_heath\\_biconf06\\_final.ppt](http://conference.ub.uni-bielefeld.de/2006/docs/presentations/cook_heath_biconf06_final.ppt) at March 1st 2006.
- Jacobsen, Dag Ingvar (2004) *Organisasjonsendringer og endringsledelse*. Bergen: Fagbokforlaget.
- Mintzberg, Henry & Westley, Frances (1992) Cycles of Organizational Change. *Strategic Management Journal*, 13, Winter, p. 39 – 59.
- Sivertssen, Svern (2006) Personal communications Feb. 21<sup>st</sup> 2006.
- Strand, Torodd (2001) *Ledelse, organisasjon og kultur*. Bergen: Fagbokforlaget.
- Van de Ven, Andrew H & Poole, Marshall Scott (1995) Explaining development and change in organizations. *Academy of Management Review*, 20 (3) July, p. 510 – 540.
- Weber, Max (1979) *Makt og byråkrati*. Oslo: Gyldendal Norsk Forlag.

# Exploitation of Informational Resources within Knowledge Society: Digital Library

By Angela Repanovici

## I. Knowledge Society

### Introduction

A new society, the knowledge society is arising, the elements of this emerging society coexisting with the constitutive elements of the industrial and postindustrial society. The most obvious aspect of the new society is the speed of use, application and dissemination of the communication and information technologies, which puts in the shade the fact that there occurs a major transformation of concepts, structures and institutions specific to the previous society. This insufficient correlation between the evolution of the information and of other components of the emerging knowledge society has led to an approach to knowledge society issues dealing with either the novelty of information and communication systems or the spiritual aspects related to it; it also maintains the confusion among the various names given to the new society, named either “post-modernist”, “post-industrial”, “informational” or “knowledge” society. The sociologist A. Giddens, considering that the new communication and information technologies have contributed to the wide world spreading of the western culture, under the form of democracy and market economy values, defines the present period as “advanced modernity” not as post-modernism. He states that modern institutions differ from all previous forms of social order only “due to their dynamism, to the extent to which they undermine the traditional habits and customs and due to their global impact”.

In fact, the reorganizations we are witness to suggest the fact that the transformations of the current period are equivalent with a radical paradigm change in the sense of the one defined by T. Kuhn in his classical work. Ever since 1993 Peter Drucker stated: “We can be sure that the world arising from the current order of values, beliefs, economic and social structures, concepts and political systems, in other words the order of conceptions of the world, will be different from what one would imagine. In some areas – and especially within the society and its structure – basic transformations have been already developed”.

[37]

The term of knowledge-society is nowadays used across the whole world. This name is an abbreviation from the structure knowledge-based society.

The cognition is the process through which there is obtained a total new side of knowledge, no matter by whom, human, animal, live machine, people network, humans and machines, and by all forms of organization that can bear knowledge. The cognitive science will get extended not only in order to integrate all these types of cognition, but also integrative, as the cognition will imply integrative processes (structural-phenomenological).

### **Knowledge, economic factor**

A new aspect of knowledge is that of economic factor. In the last 500 years, Laurence Prusak notices, the factors of production were land, labor and capital goods, being neglected the role of knowledge as distinct factor of production. To Prusak, the knowledge represents an intellectual capital, the things learnt by an organization [“there is no sustainable advantage but what a company knows, how it can use what it knows and how fast it can learn something new”].

“I don’t believe the current theory of knowledge may face these new approaches”. [31, 32]

Richard W. Everett (Chase Manhattan Bank, New York City) notices:

“Many economists have argued that technological progress is really nothing but quality improvement in human beings. Some economists take even a broader view and speak of the ‘production of knowledge’ as the clue to technological progress. The production of knowledge is a broad category including outlays on all forms of education, on basic research, and on the more applied type of research associated especially with industry ... invention and innovation.”

Roger E. Bohn remarks:

“Philosophers have analyzed the nature of knowledge for millennia; in the past half-century, cognitive and computer scientists have pursued it with increased vigor. But it has turned out that information is much easier to store, describe, and manipulate than is *knowledge*.”

This author shows that it is important to understand the **technological knowledge**, i.e. the knowledge about the best way to produce goods and services. Bohn, like others, makes a distinction between data and information.

Similarly, he makes a distinction between information and knowledge. **Data** emerge directly from measuring one or more variables. The items of information stand for data that have been organized or structured in a certain way, placed in a context and having a meaning. The **information** shows, generally or partly, the condition of the production system. **Knowledge** is much more. It pursues to understand the process, to produce causal associations, to make predictions, to make prescriptive decisions.

*It is to retain Bohn's definition for the notion of learning: "Learning is evolution of knowledge over time".*

Both define **technical knowledge** as understanding of the effects the input variables (x) have upon the output variables (Y). As  $Y = f(x)$ , the technological knowledge is the knowledge upon arguments and behavior of  $f(x)$ . The author identifies eight stages of technological knowledge. The more advanced the technological knowledge is the closer to science it is and it can, therefore, be formally managed.

The stages of technological knowledge are as follows:

- Complete ignorance of the nature of process;
- Acknowledging the process. Analogies with other processes. Enhancing knowledge outside the organization. This knowledge is tacit and placed in workers' mind. Production is more than an art. Problems are worked out through trial and error.
- Measuring stage. Variables may be measured but are yet to be controlled. This stage is preparing the following stage. It is the pre-technological stage, the knowledge is propositional, written and also unwritten.
- Variables control, but not with great precision, at a medium level. The knowledge is written or/and comprised in hardware. The learning process takes place through experiments, with scientific method. The organization is mechanically-based.
- Variables may be controlled with precision for a large range of values. There are used operation guidebooks. Workers' role changes: the focus is now on problem solving.
- Characterization and identification of processes. It is known the way variables affect the results if reduced variations of variables are produced. The process can be slightly adjusted, there can be produced reaction systems (feedback) for control. The problem solving process can be carried out through experimental scientific methods guided by appropriate theories and simulations. Learning and improving are now developing. The technological knowledge is found in

databases and software. The organization focuses on learning.

- The stage of scientific level that answers the question why. It is now outlined the scientific model of the process and the way it operates at a wide scale so that it includes non-linear and interaction effects of some variables with others. The process may be optimized and the process control may be automated. This stage is called the automation stage.

G Anthony Siesfield remarks that **knowledge cannot be measured but only its effects**. Reconsidering L. Pusak's foregoing idea, he shows that the idea of knowledge was not outlined as a stock but as a flow and in this very flow there are interwoven people's experience and inspiration leading to knowledge applied in technological processes and business management.

Ikujiro Nonaka makes some very interesting observations:

“ few managers grasp the true nature of the **knowledge-creating company**, let alone know how to manage it ... The reason: they misunderstand what knowledge is and what companies must to do to exploit it. [...] A company is not a machine but a living organism. Much like an individual, it can have a collective sense of identity and fundamental purpose. This is the organizational equivalent of self-knowledge- a shared understanding of what the company stands for, where it is going, what kind of world it wants to live in, and most important, how to make that world a reality. In this respect, the knowledge creating company is as much about ideals as it is about ideas. And that fact fuels innovation.”

In such a company the conception of advanced knowledge is not a specialized activity of the department of research and development. It is a manner of behavior, a manner of being. In such a company each one is a knowledge worker, which confers it the entrepreneur character.

Giovanni Dosi considers the economy as a distributed system of knowledge objects. This author makes the following distinction between information and knowledge:

“ The **former** entails well-stated and codified propositions about state of the world (e.g., it is raining), properties of nature (e.g., A causes B) or explicit algorithms on how to do things. On the other hand, **knowledge**, in the definition I am proposing here, includes: i) cognitive categories; ii) codes of the interpretation of information itself; iii) tacit skills; iv) problem-solving and search heuristics irreducible to well-defined algorithms.” [30]



Dosi underlines the fact that “in modern economies, firms are major, albeit by no means unique, repositories of knowledge. Individual organizations embody specific ways of solving problems that are often very difficult to duplicate in other organizations or even within the organization itself. In turn, **organizational knowledge** is stored to a large extent in the operating procedures (the routines) and the higher level rules (concerning what to do when something goes wrong or how to change lower level routines) that firms enact while handling their problem-solving tasks in the domains of production, research, marketing etc.”

Dalke Neef synthesizes the role of organizational and technological knowledge: “In the knowledge-based economy it is the production of ideas, not goods, that is the source of economic growth, and the reason that the new computing and telecommunications technologies are so economically revolutionary in their nature is that they allow ideas – in the forms of technics, research results, diagrams, drawings, protocols, project plans, chemical formulae, marketing patterns, etc, - to be distributed instantaneously and in a coherent way to anyone, anywhere in the world”.

The complexity of information not only represents the quantity of information but also makes reference to interactivity, i.e. the extent to which it can be especially elaborated for the receiver, in line with the correctness or security of information. The complexity of information means in fact the quality of information.

**Acquiring information means the number of people that can receive a certain item of information.**

The different conceptualizations of the term of informational society lay on different dichotomist sets: postindustrial society vs. informational society; mass communication vs. new communication technologies; capital goods and labor vs. knowledge, information, and communication. Beyond the variety of analytical frames, most definitions associate the informational society with a series of society-based transformations that would mark the transition to a new type of society that has become dependant on complex electronic information and on communication networks; a society allotting a significant part of its resources to information and communication-oriented activities.

Wurster and Evans notice that as long as the information is comprised within delivered physical objects, there is a basic law governing its economy: the higher the information complexity the lower the possibility to acquire it.

The new conception on knowledge capable of rendering both material and transcendental reality has been associated by some theoreticians with the technology, considered not only a discipline but also active and effective information [27]. Technology is generally described as industrially useful knowledge or as the process of acquiring this type of knowledge concretized under the form of invention, innovation, dissemination, knowledge protection, economic production, know-how, including continuities and discontinuities in information production and dissemination or the imitation and new technical solutions. However, it is considered that technology, as well as knowledge, have a transcendental dimension, a “realm of ideas”, being connected to systems of beliefs, values and preferences associated with them, which inhibits or validates certain actions not others [26].

The changes of approach to the organization’s objectives in the information era (a new form to present the new economy) as compared to the industrial era are also numerous: the mass production becomes mass customizing, the mass marketing is replaced by the individual marketing, the optimization of the information chain and the informational collaboration with providers replace the optimization of the physical chain and of the physical collaboration with providers, the virtual globalization is more important than the physical location, etc.

The new economy has generated a new type of labor, the information-based labor and a new type of worker: the information worker. He is characterized by the capacity to reason and to know at a high level of education, experience, personality and motivation, different from the worker of the industrial era.

There is prefigured the turning of the information worker into the knowledge worker. Whereas the information worker produces, processes, stores, transmits and compares information, the knowledge worker is to design information producing value, i.e. knowledge.

## **Knowledge society**

Knowledge society represents more than informational society and than computing society, interweaving the both of them.

From the moment the Internet interferes with its great advantages (e-mail, electronic commerce and electronic transactions, Internet market, “content distribution”) by comprising a great number of individuals within the electronic information area there occurs the transition to the information society.

The knowledge is meaningful meaning and active information. That is why the knowledge society cannot develop but grafted on the informational society. Likewise, it is more than informational society on account of the major role attributed to the information-knowledge in the society. The most appropriate meaning of the knowledge society is that of information and knowledge society.

The term of knowledge-society is nowadays used across the whole world. This name is an abbreviation from the structure knowledge-based society. Romano Prodi, the president of the European Commission, sometimes uses the syntagm “knowledge-based economy”.

If one is searching the theme “knowledge society” on the Internet he/she will find thousands of references. In 2001 the magazine DEUTSCHLAND dedicated a special issue to the knowledge society. Nico Stehr remarks:

“The social order shaped in the horizon is based on knowledge”

The volume of knowledge that is at our disposal doubles every five years. If we wonder about the effect of the current transition from the industrial society to the knowledge society upon the workforce and upon companies, upon politics and democracy – briefly, upon our organizational principles regarding the way we lead our life, then we are entitled to talk about the way we are going to live in the knowledge society.

The knowledge era is working [...] Knowledge is the main characteristic of the future societies. [...]

The knowledge can be defined as the capacity to act, as an action potential. The scientific and technical knowledge is nothing but the ability to take action. The privileged status of the scientific and technical knowledge in the modern society derives not from the fact that the scientific discoveries are generally considered objective, trustful, and realistic but from the fact that this form of knowledge, more than any other, continually creates new action opportunities. [...]

The scientific interpretation must reach “a conclusion” – in order to have a factual value. In the current modern society, this task of turning thoughts into conclusion and of making the scientific visions useful is due to the knowledge workers. [28]

The knowledge is becoming the basis and the principle guiding the human beings' activity. In other words, we are now organizing the reality according to the knowledge we own. [29]

If the main characteristic of the modern society is the knowledge, then the production, dissemination and design of knowledge cannot avoid the political influence. One of the most important problems we are going to be confronted with in the next decade will be the way of monitoring and controlling knowledge. This will generate the development of new branches of the science policy: knowledge policy. Knowledge policy will adjust the volume of the new knowledge that is rapidly increasing within the society and will influence its development.

This last paragraph from the above citation, which makes reference to knowledge society, recalls the notion of political technology previously introduced by the author of this study:

“The political technology is framed as a field studying the social consequences of the new technologies and is looking through the possible technologies that are likely to be achieved, in order to help the current society evolve towards a superior stage [...]. At the same time, the political technology is expected to conduct research into the consequences of the new microelectronic, computing and cybernetic technologies upon the human psychology and, inclusively, upon the society, into the changes that are produced within the structure of the labor force, into the way individuals use the time in production and in their extra productive life, as well as into an entire series of other aspects. The political technology may recommend the adaptation of the society to the new processes. [...] The political technology may formulate requirements related to technology and science in order to address the society needs, establishing a series of social functions the technical systems are to fulfill, submitting to research the way these functions may be achieved. That is why it also addresses scientists and designers of technology and technical systems. Therefore, the political technology comprises two significant aspects, one addressing the management way of the society, and the other one the innovation way. It brings together the social and technological undertakings - the way the social undertaking may favor those technologies that contribute to the economic and social progress, but also the way science and technology must gather their efforts to address the great requirements of the developing society.”

## What is knowledge society?

### The knowledge society undertakes:

- (I) An extension and elaboration of the scientific knowledge and of the truth on existence.
- (II) Use and management of knowledge framed as technological and organizational knowledge.
- (III) Generation of new technological knowledge through innovation.
- (IV) Dissemination of knowledge towards all individuals through new means, particularly the Internet and the electronic book and through new learning methods, particularly electronic means ( e-learning).

A term that has been often used recently is that of new economy. It is known that in the informational society there is framed the internet economy. In the knowledge society there is framed a new economy including the internet economy. This is the reason why the **new economy is the economy of the knowledge and informational society.**

- (V) Knowledge society represents a new economy where the innovation process (capacity to assimilate and convert the new knowledge in order to create new products and services) becomes crucial.

The innovation in the knowledge society pursues to improve the productivity, not only the classical productivities related to labor and capital goods but also the new productivities related to either natural energetic and material sources or the environment protection. That is why the new economy supposes the encouragement of creation and development of innovating plans with an own knowledge structure.

Such plans can be born through cooperation between companies, universities and governmental or public research institutes (including academic).

In a report of the European Communities Commission from 2001 it is shown that in order to get benefits from the new economy there is required a high-performance Internet interface and adequate structural reforms in society, administration and economy.

**The influence of the Internet as market in the informational society and the admittance of the assets/goods importance, especially the knowledge, stands for new characteristics of the new economy.** Richard Boulton [12]

characterizes in this way the difference between the old and new economy: in the first one it is the tangible goods that matters, in the second one it is the intangible assets creating value that are important.

The intangible is non-material, difficult to describe and especially difficult to measure.

- (VI) Knowledge society is necessary in order to assure a sustainable society from ecological point of view, since without scientific and technological knowledge it is unlike to produce goods, organizations and economic and technological (even biological) changes necessary to save the humanity from disaster in the XXI-th century, Consequently, the knowledge society identifies with the informational and sustainable society. It will be very difficult to find another way for sustainability, beside the knowledge society.
- (VII) Knowledge society has global character and stands for a globalization factor. Through both components, informational and sustainability, the knowledge society will have a global character. The knowledge itself, as well as the information, will have a global character.
- (VIII) Knowledge society will also stand for a new stage in culture, the knowledge culture passing in the forefront, implying every type of knowledge, the artistic, literary etc knowledge inclusively. This way, there will be paved the way for what we have called Society of conscience, of truth, morality and spirit.

In table 1.1 the main features of knowledge society will be synthesized:

|   |
|---|
| Knowledge society presupposes:  |
| I. An expansion and a thorough study of the scientific knowledge and of the truth about existence   |
| II. Use and management of the existent knowledge in guise of technological and organizational knowledge   |
| III. Production of technological knowledge by means of innovation   |
| IV. An unprecedented dissemination of knowledge towards all citizens by new means, resorting with priority to the Internet and to the electronic book, and resorting to methods of learning by electronic procedures (e-learning) |

|  |
|--|
| V. Knowledge society stands for a new economy in which the innovation process will play a decisive role. The influence of the Internet as market in informational society and the acknowledgment of the importance of the intangible assets value, especially the knowledge, constitute characteristic features of this new economy. |
| VI. Knowledge society is fundamentally necessary in order to ensure an ecologically sustainable society.   |
| VII. Knowledge society has a global character and is a globalization factor.   |
| VIII. Knowledge society stands for a new stage of culture.   |

**Table 1.1 Components of knowledge society**

If informational society is applied the vectors of knowledge society, even during the development of informational society, it is then possible to gain time. Therefore, in our country, the actions undertaken for knowledge society should unfold simultaneously with those concerned with the passage from the informational underdevelopment stage to the informational development. Not in the first place informational society and afterwards knowledge society, but the pursuit of both objectives should be combined from the very beginning.

This is the only way for us to reach a developed informational and knowledge society.

### **Vectors of Knowledge Society**

Two comprehensive classes of knowledge society vectors have been defined:

- Technological vectors
- Functional vectors

A vector of knowledge society stands for an instrument that transmutes informational society into a knowledge society. In order to make the first move in knowledge society, it is necessary to launch a minimum number of such vectors. The first such vector consists in the creation of a “developed” Internet, which is a technological vector, then *the technology of the electronic book* (technological vector) and the knowledge management (functional vector, endowed with two valences, the first for the economic and organizational functioning of a multinational enterprise, corporation or of society itself, the second for the moral use of knowledge within global society). However, the number of these vectors of knowledge society is much more numerous, every new vector making a step forward for the development of this society.

## Technological Vectors of Knowledge Society

We deem the following technological factors to be taken into consideration for knowledge society:

**Developed Internet**, by geographic expansion, by use of transmission bands up to the largest possible, by passing from the IP4 communication protocol to the IP6 protocol, by including each and every institution, residence and citizen within the network.

**Technology of electronic book**, which differs from the book on the Internet, although its diffusion is based particularly upon the Internet, but also upon CDs. **Intelligent agents**, which stand for expert systems endowed with artificial intelligence, used for data mining and even for knowledge discovery; the intelligent agents will be made much use of for many functional vectors of knowledge society.

**Intelligent environment** for human life and activity. **Nanoelectronics**, which will turn into the main physical support for processing the information, as well as for many other functions, belonging both to knowledge society and to conscience society.

## Functional Vectors of Knowledge Society

The number of these factors may be great enough, as more and more fields of activity increasingly depend upon knowledge:

- Knowledge management for national and local enterprises, organizations, institutions, administrations.
- Management of moral use of knowledge at a global level.
- Biological, genome knowledge (knowledge of the genome and of the functions the genes structure determines).
- Health care system at a social and individual level.
- Environment protection and ensuring of durable and sustainable society by a specific knowledge management.
- Thorough study of the knowledge upon existence.
- Generation of new technological knowledge.
- Development of a culture of knowledge and innovation.
- An educational system based upon the methods of informational society and of knowledge (e-learning) etc.



|  |
|--|
| <b>General theory of information</b>   |
| <b>Information</b>   |
| Phenomenological   |
| Structural-phenomenological<br>- Integrative<br>- Mental                                 |
| Structural<br>- Syntactical<br>- semantic (information with its own, internal meaning)   |
| <b>Data</b>  |
| Syntactical structural information with an external meaning, in a certain context        |
| <b>Computing program</b>   |
| Information which acts as syntactical information or as intelligent semantic information |

**Table 2: General theory of information**

## **Knowledge Management at Global Level**

The greatest danger which stands out for knowledge society resides in the utmost extension of knowledge society privatization. An equilibrium must be made up between the economic and the moral use of knowledge. There must be defined the sphere of management for the moral use of scientific knowledge within global society.

The knowledge management at a global level should aim at one of the basic targets of knowledge society, that is the dissemination of information-scientific knowledge on the largest scale possible, free of charge or at an extremely low price. This tendency is checked by another contrary tendency, imposed by the new rules of intellectual property introduced especially during the last 10-15 years.

We may take into consideration the following typology:

- Technological and economic systems (global structures; restructuring and economic networks; state-of-art information and communication technologies; new practices of media communication);
- Institutional changes (societal networks; crisis of modern institutions; decline of communities and of traditional social relations; emergence of new types of community; a new public space);
- Ideological and valor changes (neo-individualism; new cultural values and changes; postmodernist practices in everyday life).

## II. Role of the Library – Infodocumentary Institution within Knowledge Society

In the framework of the general direction, informational society of the European Commission, one of the priorities aimed at consists of the amplification of the functioning efficiency of the great cultural effects depositaries by means of state-of-art management and interfacing techniques.

### Definitions of the Digital Library

- “New digital libraries are characterized by features which are not possible for traditional libraries, this way the concept of library is expanding considerably beyond the physical limits. They will provide innovative resources and services. An example consists in the ability to interact with the information: rather than placing a reader in front of a table with numbers, digital libraries allow users to perform a selection among various ways of visualizing and operating with the numbers, including the graphic charts they can explore. By means of an extensive use of hypertext links in order to interconnect information, digital libraries allow users to find related digital materials upon a certain subject.” [127]
- “Digital libraries stand for organizations which provide the resources, including the specialized staff, for selecting, structuring, offering intellectual access to, interpreting, disseminating, preserving the integrity and assuring the duration in time of the collections of digital works so that these ones should be promptly and economically available for use by a pre-established community or by a set of communities.” (Digital Library Federation)
- “Digital libraries represent complex data/information/knowledge (henceforth information) systems which contribute to: the compliance with the information necessities of the users (societies), the supply of information services (scenarios), the organization of information upon useful ways (structures), the administration of the information location (spaces) and the settlement of a communication channel with the users and their agents (channels)” (Edward A. Fox, July 1999, in conformity with 5S Framework)
- The activity of the digital library unfolds in the framework of a complex study location molded by four dimensions: community, technology, services and content” [99]
- “The field of digital libraries deals with the amplification of human civilization by applying digital technology to information problems addressed by institutions such as: libraries, archives, museums, schools, publishing houses and other information agencies. The activity upon digital libraries focuses

upon the integration of the services and upon a better compliance with human necessities, by means of a holistic treatment regardless of interface, location, time, language and system. Although substantial creations have been created for the individuals' exclusive use, we deem shared resources one of the defining characteristic features of libraries. Libraries perform connections between people and information; digital libraries amplify and augment these connections" [100]

- For a conclusive discussion upon the community definitions, approaches and perspectives upon "digital libraries", you may refer to "What do digital Libraries represent? Visions in Competition" [9]
- "The digital library is
  - The collection of services
  - And the collection of information objects
  - Which assist the users in the operation with the information objects
  - And the organization and presentation of the respective objects
  - Available directly or indirectly
  - Via electronic/digital means." [ 32]
- "The digital library stands for a concept which displays different meanings for different communities. As for the engineering and computing community, the digital library is a metaphor for the recently distributed types of data bases services which manage the non-structured multimedia data. As for the political and business communities, the concept represents a new market place for the world informational resources and services. As for the futurist communities, digital libraries stand for the Well's World Brain manifestation. The perspective taken into consideration is rooted in the tradition of information science." [99]
- "An organized data base, comprising digital information objects in different formats, maintained in order to supply unmediated access to a community of users, which displays the following characteristic features:
  - a comprehensive access means (for example a catalogue) has at its disposal a research and retrieval capacity within the entire data base;
  - there are organized technical procedures by whose means the library management adds and removes objects to/from the data base in conformity with a coherent and accessible policy of the collections." [31]
- "Digital libraries constitute a set of electronic resources and technical abilities associated to the former for the creation, search and use of information. In this respect they stand for an extension and for an amplification of information depositing and of retrieval systems which manipulate digital data in any medium (text, images, sounds; static or dynamic images) and which are present within distributed networks. The content of digital libraries includes data, metadata which describe various aspects of the data (for example representation, creator, possessor, copyrights) and metadata which consist of links or relations with other data or metadata, either internal or external to

the digital library (UCLA-NSF Social Aspects of the Workshops of Digital Library).

- “Digital libraries are built – collected and organized – by a community of users, and their functional abilities support the information necessities and the manners of making use of the information which characterize the respective community. They constitute a component of the communities in which individuals and groups interact with each other, resorting to data, information and resources and knowledge systems. In this respect, they constitute an extension, amplification and integration of a wide range of information institutions, as physical places in which resources are selected, collected, organized, preserved and accessed for the support of a community of users. These information institutions comprise, among others, libraries, museums, archives and schools, however digital libraries extend and serve other community groups, including classrooms, laboratories, hostels and public spaces.” (UCLA-NSF Social Aspects of the Workshops of Digital Library)”
- “Systems providing a community of users a coherent access to a vast, organized thesaurus, of information and knowledge. This organization of the information is characterized by the absence of a previous detailed knowledge of the ways of making use of the information. The user’s ability to access, reorganize and make use of this thesaurus is enriched by the capacities of digital technology” (adaptation from Interoperability, Evaluation and Research Agenda of Digital Libraries)
- “A library that has been expanded and enhanced by the application of digital technology. Important aspects of the digital library which may be expanded and enhanced include:
  - Collections of the library
  - Organization and management of collections
  - Accession of the library units and processing of the information comprised in these units
  - Communication of the information with respect to these units
  - “The generic name for the federate structures that offer humans both intellectual and physical access to the enormous and in continuous growth networks of information encoded in multimedia digital formats.” (Digital Library of the University of Michigan: It’s not Your Father’s Library – Bill Birmingham)
- “A digital library constitutes a medium of distributed technology which dramatically reduces the boundaries from the creation, dissemination, manipulation, depositing, integration and reuse of the information by individuals and groups.” (Edward A. Fox, editor, Source Book upon Digital Libraries)
- “A digital library is a representation readable by a machine of the materials, which can be found in the framework of a university library, together with the organization of the information meant to assist the users who search for

specific information. A service of the digital library stands for an assembly of digital processing, depositing and communication machinery, together with the software necessary to resume, emulate and expand services offered by conventional libraries based upon paper and other material means for collecting, depositing, cataloguing, retrieving and disseminating the information (Edward A. Fox, editor, Source Book upon Digital Libraries)

## **Glossary/Terms related to Digital Library**

(by Peter Graham, Rutgers University Libraries)

- digital archive: a digital library which is intended to be maintained for a long period of time, i.e. longer than an individual's life span and surely longer than the individual technological eras (previously, sometimes, also "digital research library")
- digital preservation: preservation of the artifactual information by the digitization of its image (for example scanning a manuscript page, digitally photographing a vase, or converting the registration of a cylinder under a digital form).
- electronic preservation: preservation of the information which exist under digital form (which means electronic), i.e. techniques associated to refreshing, migrating and assuring the integrity.

## **Techniques of Digital Preservation**

- Refreshment: copying the digital information from a long-lasting depositing medium in another medium of the same type, with not the least change in the bit flow (for example, from a degraded 800 bpi band to a new 800 bpi band, or from an older 5" floppy to a new 5" floppy)
- "Modified refreshment" means copying in another medium of a partial similarity so that no change should be operated in the bit pattern that is connected to the application and operation system which makes use of the data, for example from a 800 bpi band to a 1600 bpi "square" band with cartouche; from a 5" to a 3" floppy disk.
- Migration: copying data or converting data, from a technology to another, either hardware or software, maintaining the basic characteristics of the data; as a matter of fact far in the future. (Nowadays, it is known, this final qualifier raises numerous questions.) Example: conversion of XyWrite w/p files in Microsoft Word; conversion of ClarisWorks v3 spreadsheet files in Microsoft Excel v4 files; conversion of binary band images of research multi-punched cards in database format; copying of a 800 bpi band file on a sequential disk file; conversion of a database in DOS Fox Pro in a database Visual Basic for

Windows 95; conversion of a PICT image in a TIFF image; conversion of a ClarisWorks file for Windows v4 w/p in a file Macintosh ClarisWorks v4.

We may provide examples, as those offered above, for cases we know to be required; the issue of the long term preservation consists in adopting measures for subsequent migrations, not knowing what the future technologies will be made of.

- Emulate: from the point of view of the hardware terminology, creation of software for a computer which reproduces in all essential characteristics (as they are traced through the intermediary of the problem which is to be solved) the performance of another computer of a different design. Computers may emulate computers from a previous generation in order to ensure the compatibility, or they may emulate a computer from the generations to come in order to offer a media for the development of the software, while the new computer is still in the manufacture phase.

From the point of view of the preservation terminology, the creation of a software which analyses the software media of a document in such a manner that it should provide an interface between the user and the document which reproduces in a significant manner the characteristic features of the document the way it was created from the original software.

### **III Exploiting the Digital Library Resources in the Knowledge Society**

The digital library provides services based on the new information technologies. The users wish to find information within the shortest time possible.

The collections management, funds designated to acquisitions are oriented towards data bases and on-line services.

In order to evaluate services and to improve the performance indicators within university libraries, managers are confronted with similar problems:

- Types of users accessing the services provided;
- Collections accessed;
- Ways to help digital library services become more efficient when meeting users' requirements.

The library users are interested in: the way in which users access these digital libraries and, due to the environment nature, they can access detailed recordings with regard to the way users navigate through these informational spaces. In order to exploit and manage great amounts of information in the digital library there is advanced an approach to these data bases through data mining applied in libraries.

## **Data mining (DM) – general outlines**

**Data mining (DM), known as knowledge-discovery in data bases (KDD)**, is the process of automatically searching large volumes of data for patterns (Wikipedia article “*Data mining*”<sup>1</sup>). In order to achieve this, data mining uses statistics and pattern recognition.

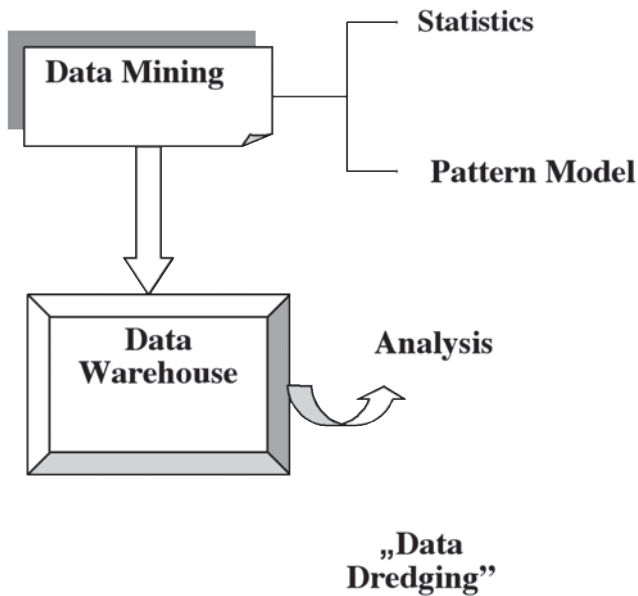
**Data mining** was defined as being “the nontrivial extraction of implicit, previously unknown, and potentially useful information from data” but also as being “the science of extracting useful information from large data sets or data bases”.

Data mining is a general term and it is used at a large scale of contexts with various meanings.

Used in the technical context of data warehousing and analysis, data mining is a neutral term. However, the term was sometimes used with pejorative meaning, imposing patterns (and, in particular, causal relations) on data, where they did not actually exist. This imposing of non-relevant correlations, which induce errors or attribute-based trivial correlations, is the criticized term from “data dredging” statistics. In a more constraint sense “data dredging” implies data scanning for any relations and, when something is found, there is given an interesting explanation. The problem consists of the fact that, invariably, within the masses of data there arise interesting particular relations. Another danger consists in discovering correlations that do not usually exist. The investments analysts are the most vulnerable ones in this area.

---

<sup>1</sup> [http://en.wikipedia.org/wiki/Data\\_mining](http://en.wikipedia.org/wiki/Data_mining)



**Figure 1 : Elements of data mining**

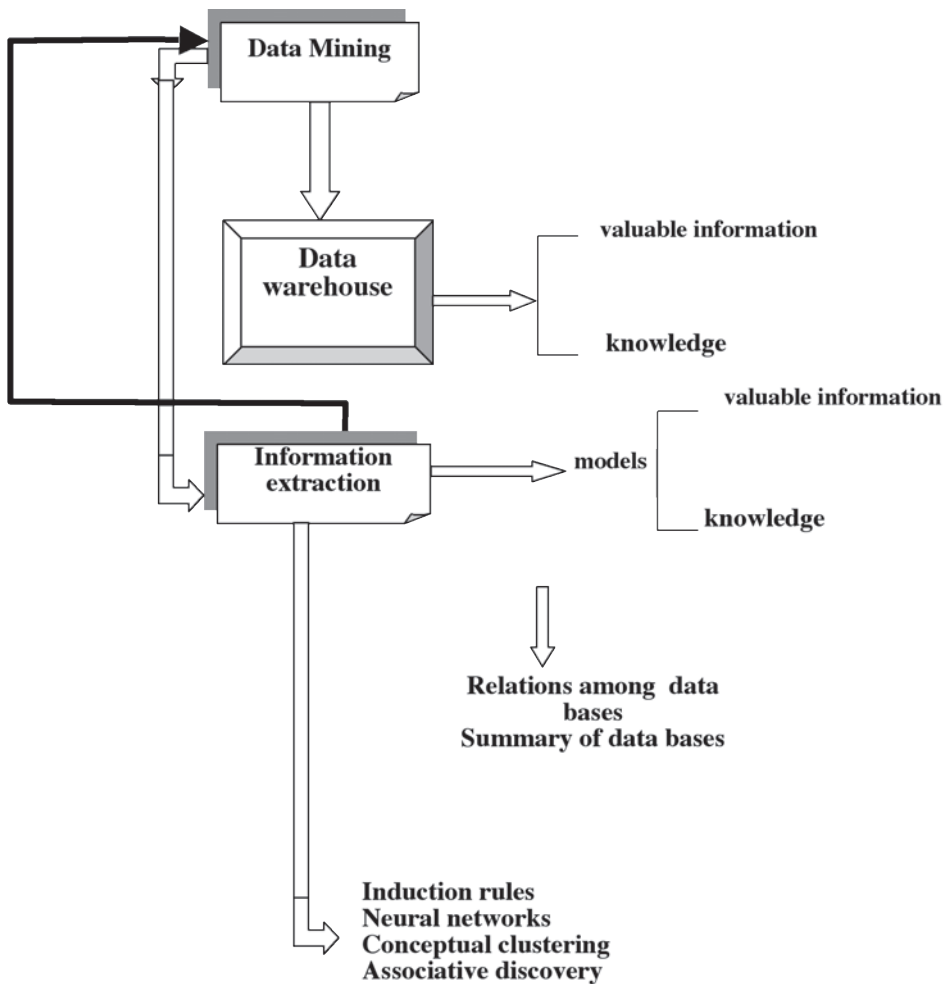
In **data mining** there was made great effort to develop a model of fine granularity and as detailed as possible for data masses. In *“Data Mining for Very Busy People”*, the researchers from West Virginia University and University of British Columbia examined a method that involves minimal differences among the elements of a given amount of data, with the aim of developing simple models representing relevant data.

In essence, **data mining** gives information that would not be available otherwise. In order to be useful they must be correctly interpreted. When collected data involve other persons there occur more problems related to confidentiality, intimacy, legality and ethics.

**Data mining** consists of a multitude of techniques in continuous development that can be used in order to extract valuable information and knowledge from large volumes of data. Researches into data mining and related instruments would put the accent on commercial applications. Few researches made reference to scientific and satellite data. Despite the conferences’ debates on various aspects related to data mining and scientific **data mining**, an exchange of ideas focused on **scientific data mining has not really occurred between scientists and specialists in data mining.**



Data mining is a stage in the process of knowledge discovery consisting in applying data analysis and discovery algorithms that, within reasonable limits of calculation efficiency, produces a particular enumeration of data patterns. It is to mention that the methods, the number of variables considered may be reduced or there can be found invariant representations for data.



**Figure 2 : Data Mining elements and links**

In compliance with the Charta of data management solutions provided by IBM, data mining is the process of extracting valid, previously unknown, and intelligible information from a large data base, information used to make decisive decisions in the business field. Information extraction may be used to shape classification or prediction models, to identify relations among the data bases records or to provide a summary of data bases that are mined. Data mining consists of a certain number of operations, each of them being supported by a variety of techniques such as induction rules, neuronal networks, conceptual clustering, associative analysis, etc. In many fields from the real world such as marketing analysis, financial analysis, fraud detecting etc., the information extracted asks for the cooperative usage of more data mining operations and techniques.

Consequently, data mining is an area undertaking continuous development in the field of computers science, which will provide a new and efficient level of information and knowledge discovery that all users in the field of computerized data storage will benefit from.

## **Data Mining, a new era in computing**

### **Knowledge exploring**

“... numerical information is easily acquired and cheaply stored. But what could be done with such a great amount of data?”

In order to make decisions, as revealed above, human beings need knowledge. Therefore, there emerges the issue of data analysis and knowledge extraction from them.

If data are relatively few in number, could either specialists in various fields or statisticians easily and manually fulfill the analysis? Some authors call them “miners” or “manual excavators of data”. What are we going to do if we have great amounts of data? How large could a database be in an application? An answer could be found in the Case study box: University Library.

A new discipline was born under several denominations, Data Mining, Knowledge Discovery (KD), Knowledge Discovery in Data bases (KDD), Information Discovery (ID), Information Archeology (IA) etc. Each denomination

may be justified in its own way. Yet, we have to underline that some authors trace differences among them, for example between DM and KDD or KD.

Therefore, what is DM? From the great amount of definitions we choose only two that seem more suggestive and simple. “DM is extracting predictable information from large data bases” or “torturing the data until they confess”.

## DM and KDD components

The main function of DM is, therefore, to extract patterns of knowledge from data. For this very reason, DM makes use of a variety of statistics algorithms, forms recognition, classification, fuzzy logic, machine learning, genetic algorithms, neuronal networks, data visualization, etc. The variety of algorithms may be grouped in the main components of DM. The number of these components differs from one author to another. Therefore, some authors consider that DM has 3 components, others consider that DM 4 etc. We consider that the main components of the DM are:

- *model* – represented, like any computing model, through a function in a one-dimensional or multi-dimensional space (an assembly of functions), depending on parameters. It can be represented as either a linear function of parameters or a probability function or fuzzy function etc. The model is conceived with several algorithms, such as classification and clustering algorithms;
- *preference criteria* – can be of different nature, some of them being based on ordering, others on interpolation or on the best approximation;
- *selection algorithms* – leading to a selection of three important elements occurring in DM, that is: the model, which is selected from the models basis, data, which are selected from the database and represent parameters, and the criterion or criteria of preferences, which is selected from the criteria base;
- *setting the deviations* – generally consisting of algorithms determining the deviation or the stability; a specific category of such algorithms are the statistical ones, through which there are established the model’s deviations as compared to the ideal one.

The authors that make the difference between DM and KDD consider KDD as being an iterative and interactive process that includes DM. Therefore, as part of KDD the knowledge extraction is accomplished by following the steps:

- *learning the application-related field* – consisting in the acquisition of knowledge about the initial condition as well as about the particular goal of the application;
- *establishing the data set* – to which is applied the process of knowledge extraction; most authors draw the attention upon this stage called the preprocessing stage;
- *cleaning and selecting data* – which is a complex process, specific to data warehouses, coping with noise elimination, dimension diminishing, data replacing, etc;
- *applying the DM procedures* – which is in fact the most important step of KDD process;
- *interpreting results* from user's point of view – which is a decision-making stage; if he is not satisfied with the result, he can go back to any of the previous stages;
- *using the knowledge discovered* – which is the final stage; this use is accomplished by incorporating them within integrated knowledge systems or by concluding simple reports to those interested.

It is obvious that the specialists considering KD or KDD synonyms with DM, state that these stages are part of DM stages.

### **Link between DM and other fields of computing**

DM and KDD are connected to new fields of computing. Our intention is not to present all these connections. However, we underline that the closest fields to DM and KDD are OLAP (On Line Analytic Processing) and DSS (Decision Support Systems). There are many presentations, more or less complete, on OLAP and DSS. According to them, OLAP is a way to use data warehouses, usage that supposes, on the one hand, on line access (OLTP – On Line Transactional Processing) and, on the other hand, a multidimensional (vector-based) analysis of large data bases. DSS is an assembly made up of data bases and data warehouses, as well as other assemblies of useful information, together with soft products conceived to draw up reports, data analyses and to implement optimization algorithms in view of supporting the process of decision-making undertaken by business people. Since these two fields are strongly connected to DM they determine a series of OLAP or DSS providers to sustain the DV delivery. “Giving a weak interpretation to the notion of Data Mining, the OLAP providers may say they deal with DM” states Steve Smith, director of advanced analyses at Pilot Software. “One of the

lines of demarcation between the two fields is the automation degree. The second one is the degree of using the statistical means to determine what is interesting and relevant”.

These are definitely in favor of DSS. OLAP and DSS use fewer algorithms and depend on the requests and hypothesis delivered by the user. As far as the answers are concerned, the difference between OLAP and DSS, on the one hand, and between DM and KDD, on the other hand, may be associated with the difference between the answers given by a data knowledge and bases. Consequently, for instance, a typical way to formulate a question in OLAP and DSS is: “*Did users from Norway access more about vehicles as compared to the users from Romania last year?*”. The OLAP and DSS answer takes the following form: “*The Norwegians accessed 12 000 publications about cars, whereas, the Romanians have accessed 10 000 within the same period*”. This answer is therefore clear but limited to hypothesis and, therefore, rigid. The answers in DM and KDD are much more flexible. In their case, the questions take the following form: “*Present a model that identifies the most predictable characteristic of the population that is to access publications about cars*”. Learning from past experience, DM will answer: “*It depends on period and department. Therefore, for example, last year users from Norway accessed more publications about cars (12 000) than those from Romania (10 000)*”. The advantage is that DM and KDD may discover more sophisticated and subtle answers that OLAP and DSS cannot detect.

There are even more differences, among which one of the most important is the reasoning way.

Therefore, the reasoning in OLAP and DSS systems is deductive, while in DM and KDD it is inductive.

On the other hand, it must be underlined that the differences between DM, OLAP and DSS have the tendency to disappear. The DM algorithms and facilities have been more and more often introduced in DSS and OLAP products. For example, AT & T, which is a great OLAP user and provider, has started to work on a new product DM, named Nearest Neighbor System that classifies the users according to similarities. At their turn, the DSS providers have specific visualization systems of data bases; therefore, introducing some DM functionalities is a normal option. It is not to wonder that the OLAP and DSS providers, such as Pilot and Comshare have rapidly oriented towards DM.

One of the favorite fields of DM usage, by OLAP and DSS and producers is Internet enhancement. This fact appears as natural, viewing the heterogeneous character of data bases and warehouses in this case. Therefore, for instance, Tree Software has introduced a Web gateway at a price of 10.000 dollars, a gateway that is a site between the Web server and a multidimensional OLAP database of type Essbase of the company. The product assures the writing and reading of

business information analysis furnished by the browser Web, which will facilitate its spread on DSS, OLAP and DM markets.

## **Data Mining applications in digital library – Bibliomining**

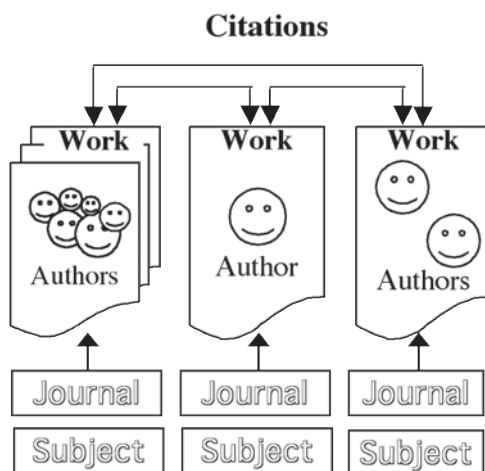
### **Origins and definition of bibliomining – Theoretical Concepts**

Bibliomining derives from the terms “bibliometry” and “data mining”, since the aim is to benefit from social networks that justify the functioning of both bibliometry and user-oriented data mining, through the intermediary of a unique center of data storage. Why should a new term be created for data mining in libraries? The concept is not new; data mining has been occasionally analyzed within library science meetings beginning with the last half of the 1990 [5]. The challenge resides in the terminology used; the packages of data mining comprise a library of various algorithms. That is why searching for articles in a particular field may lead to documents that are not on topic, such as “Re-usage patterns aiming at data mining in Library for the applications selected by user” [108]. In view of facilitating the activity of specialists in libraries-oriented data mining and in view of avoiding the confusions of library-oriented software for data mining, Nicholson created the term of “bibliomining” for the paper from 2003 by Nicholson and Stanton.

In order to better conceptualize the bibliomining concept it is useful to first conceptualize the data required for traditional bibliometry and user-oriented data mining and then to see how they can be combined to create the bibliomining concept.

### **Bibliometry**

Bibliometry is based on quantitative exploitation of document-oriented scholastic communication [9]. Figure 3 presents a part of the data used in bibliometric research and the connections among different papers. The papers have authors and collections (periodicals, publishing houses, libraries) associated to them, and the papers are interconnected by the intermediary of references, work paternity, common terms or other aspects of the conception and publishing process.



**Figure 3: Connections among the papers used in bibliometric research**

The traditional bibliometric data imply information regarding the paper's conception, such as paternity and cited works. In addition, the metadata associated to the paper, such as the general thematic or particular periodical where it occurred are connected to data viewing the paper's conception. The association of these data enables the researcher to understand the context in which the paper was conceived, the impact of citation on long term and the differences among fields as regards their patterns of scholastic results.

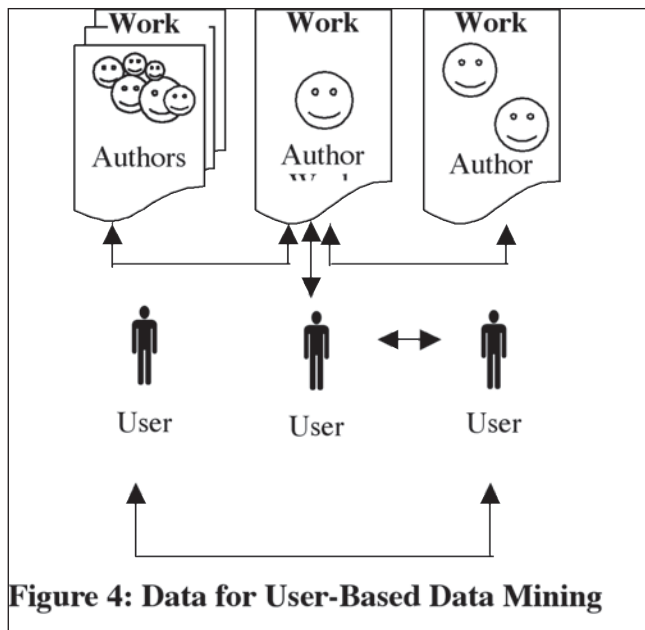
The analyses carried out in traditional bibliometry were based on frequency; still, many recent bibliometric studies resort to visualization and data mining in order to explore the patterns by creating these materials [90; 11]. A part of the concepts explored includes the paternity frequency in a field, the generality degree of words used and the discovery of a central set of frequently cited works [10]. The reference integration among works allows a rich exploitation of relations between scientists and subjects tackled; the connections among papers are used for automated searching of information and erudition visualization [155] and of social networks [136] for those involved in the creation process. Many recent bibliometric applications involve Web-based resources and hyperlinks that emphasize or substitute traditional information linked by references [9].

### **User-oriented data mining**

A popular area of data mining not only in the commercial sector but also in the scholastic literature consists in investigating the way users explore Web spaces. These studies are centered on Web pages accession by a certain user (or IP

address). There are discovered usage patterns through data mining that are used to personalize information offered to the user or to improve the information service [156; 38]. Figure 4 reveals a part of the data used in this user-oriented data mining. The aim of this figure is to demonstrate that in user-oriented data mining the links among papers come from a common usage. If, for example, a user accesses two papers during the same session, when another user visualizes one of these papers, the other might as well arise his interest. This figure presents the links among works that result from users.

A record in this data source represents a unique accession of the data resource, and the metadata attached to this recording include any available



identification as regards the user, information related to time and datum and data related to the corresponding Web site. Certain studies append metadata with regard to the work accessed in view of improving the recommendation algorithms [51; 87]. The patterns are oriented towards the understanding of the manner in which the users explore the informational space; if there is any way to identify a user between sessions (through a cookie or a login), the users' behavior may be detected throughout time. Since many digital library services require a certain type of login to access materials bought or rented, this type of data mining in use is possible and will be useful in the decision making process.

The challenge of implementing this type of exploitation in developing the activity of a digital library consists in that of user intimacy. The private character



of information making reference to the user that can be identified as person represent a real preoccupation during the bibliomining process.

A commonly inferred solution consists in coding the user's identification data in the center of data storage. This would allow a detection of articles within a period of time and it would not allow an immediate identification of users in the center of data storage. Yet, the substitution of the user's ID with a code is not appropriate as anyone wishing to find out information regarding the user's behavior may resort to the codification diagram to find the codified ID for a user and to find the information regarding the user. That is why the codification may represent a tempting choice but if the codification procedure is reversible or a codification study is maintained, the user's private character still represents a worry.

### Data storage center for bibliomining

Both bibliometrics and bibliomining for Web usage have a data field in common – the work accessed. On the one hand, the bibliometrics is focused on work's creation, and, on the other hand, the exploitation for Web usage is focused on work's accession. The combination of these two sources of data in a single center of data storage allows the researchers and library managers to fully understand the information space created. Figure 5 presents the model for data resources in bibliomining, which demonstrates the concept of connections among works on the basis of both the creation process and users' population.

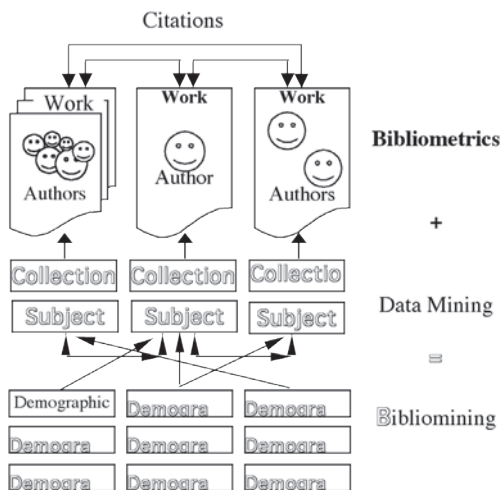


Figure 5: Data for Biblio-Mining

The bibliomining is defined as combination between data mining, bibliometrics, statistics and reporting instruments used to extract artifacts patterns based on behavior in library systems [113]. It was rarely implemented in full form, as it was presented here, due to the complexity of problems related to data. By presenting the model the goal is for the institutions developing centers of data storage for digital libraries to keep in mind the complete center of data storage in bibliomining as aim while developing smaller projects. The integration of bibliomining within current research and evaluation will allow managers and researchers in the library to make a more complete idea upon the resources held by their library and on the way they are accessed by users.

### **A framework for data**

The data that can support the links between both creation/publication and usage have to be found in the same center of data storage in view of allowing the whole bibliomining process to develop. A conceptual frame for these types of data is useful to determine the fields that are better to be excluded from the operational systems. There are three parts making up this frame - data regarding the work, data regarding the user and data regarding the service. These three parts will be interwoven to represent a usage that is at the basic level for the center of data storage. The case of using a library service connects one work (or works), a service and a user in the center of data storage for bibliomining.

- I. The first section of the center of data storage comes from **work**. This will comprise three types of fields – fields that were extracted from the work (like title or author), fields that were created as related to the work (like subject heading) and fields indicating the format and work's location (like URL or collection). These items of information may come from a MARC recording, Dublin Core information or from the management system of library content. This area may be converted into bibliometric information, such as references or connections with other works. This may ask for the extraction from the original source (in case of digital references) or connection with a database for references. A challenge of creating this link consists in the fact that providers currently report a use of electronic resources at the title level, while many bibliometric explorations begin at the level of article. There are required standards for reporting at the level of article linked to COUNTER formats; once created, the providers may provide more information regarding the specific elements used at their sites.
- II. The second area of the center of data storage comprises items of information regarding **the user**. As previously mentioned, this is the place where the

demographic surrogate is to be stored. Supplementary, there will be stored other fields that come from user interferences. For instance, the user's IP address may activate usage place interferences. In the case of university libraries the IP address could be used to deduce if someone comes from off-campus, on-campus or from the library. In certain cases the IP address could provide information regarding the building or the computing laboratory where the request comes from. A similar inference from the public library may come from the zip code. Both inferential connections may provide demographic keys with regard to the users' groups, but neither will provide a complete demographic superposition. However, at an ample set of data there may be deduced patterns from these best demographic keys.

III. The third area comes from **the main reasoning of library utility** – to connect users to information, especially by the intermediary of **works**. The most difficult part to be conceptualized is the library service since there are many and various types of services. **Searching, circulation, references, inter-library loan** and other library-oriented services have in common fields that may be captured in the center of data storage. Supplementary, each of them presents a set of appropriate fields for each type of service. A center of data storage adequately conceived may handle both types of data; this situation allows the evaluation of a certain type of service or the acknowledgment of library utility by the intermediary of its multiple services. The fields common to most services include time and data, library staff involved, location, method and if the service undertaking was made in association with other services.

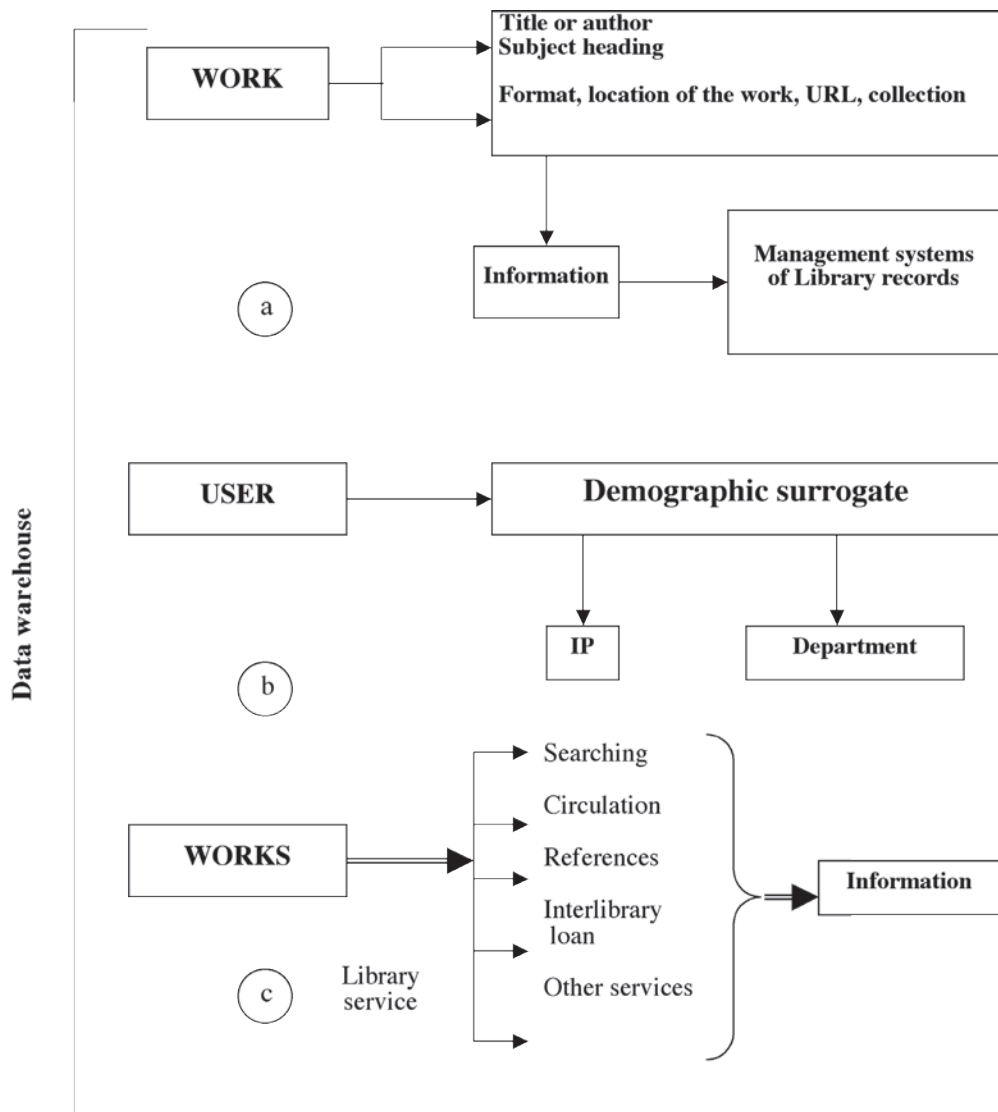
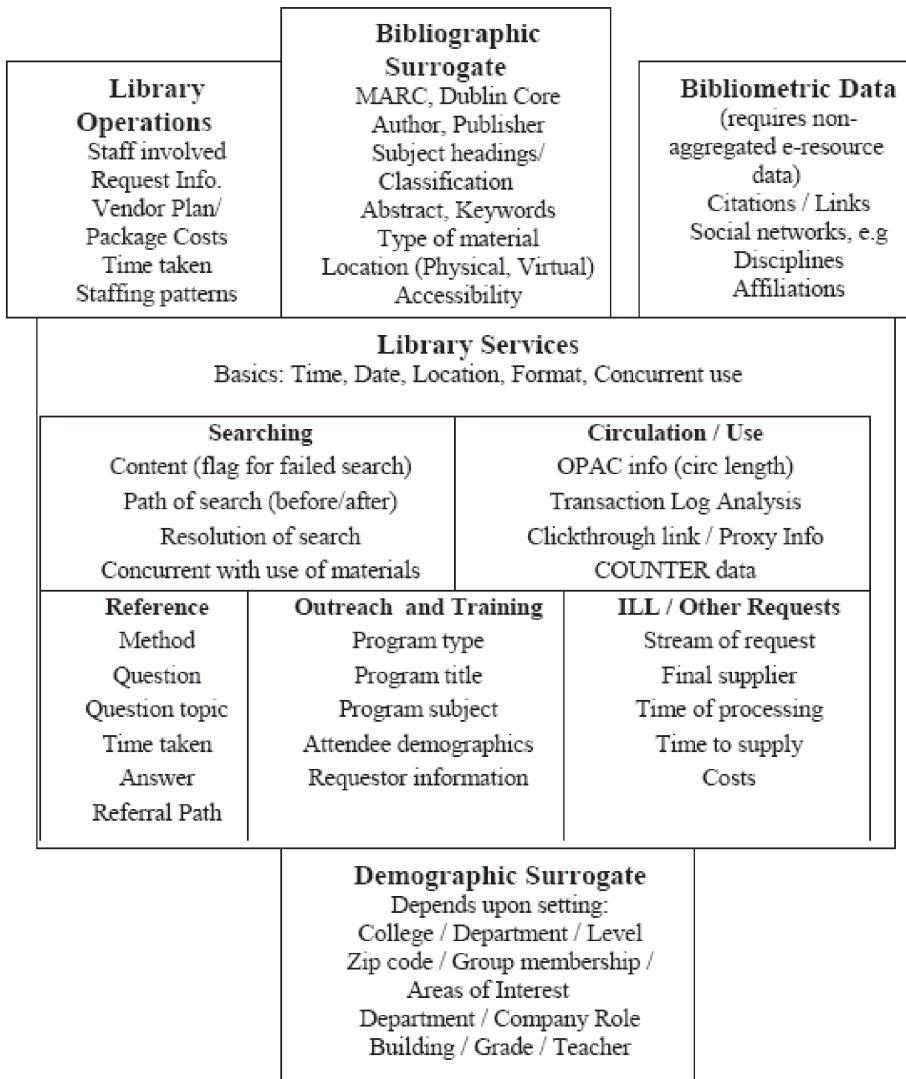


Figure 6

Each library service presents a set of adequate fields. For instance, the searching presents the searching content and the following steps to be undertaken. The inter-library loan presents the cost, a provider and an interval to fulfill the request; the circulation presents information with regard to the acquisition of work and duration of circulation. As regards most of the decisions, the necessity to make all the decisions assigned to the library and the necessity for scientists'

research should guide the captured fields while still keeping the user's intimacy. In order to support the exploitation, Figure 7 comprises many other components and fields that demonstrate the conceptual frame for the storage center of data within bibliomining.



**Figure 7: Conceptual framework for data types from data warehouse**

## **Using the center of data storage for bibliomining**

Once the source of data is collected, it allows not only the creation of traditional reports and measures but also the creation of new exploitation opportunities. This section presents various instruments that are useful in exploiting the center of data storage; the goal of this section does not consist in explaining the way to use each instrument but in presenting the instruments within the context of structure of the center of data storage for bibliomining and the way in which these instruments could be further developed and implemented.

### **Traditional reporting**

Traditionally, the persons responsible for making decisions in a library investigate the media in order to understand the way their service is operating. These measures may be all created together with this center of data storage but, at the same time, they present the advantage of enabling the managers and persons in charge of making decisions to ask other questions. If there are collected amounts of data exclusively and the subjacent data are eliminated (or are never put at disposal), then the ability to ask new questions is reduced to a great extent. The questions must be formulated in time and once the reports are drawn up they are difficult to modify as it is likely that the measures are not comparable in time. An exclusive argument to keep the amounts of data consists in the fact that it protects users' intimacy; despite all these, the utility of demographic surrogates allows the library to keep supplementary data with regard to its users and services while protecting their intimacy.

The advantage for the center of data storage consists in the fact that new questions may be formulated not only with regard to the current situation but also as far as the past is concerned. Since the center of data storage represents a collection of past actions, from past actions there may be collected new media and amounts of data. This allows the persons in charge of evaluation and measurement to think up new questions and then to look through these reports in view of understanding the tendencies. It would be an impossible task if only the amounts of data were kept; there could be formulated exclusively new questions and there would not exist any background to implement measures in perspective.

Furthermore, libraries may understand more easily the behavior of different demographic groups within the library. The amounts cover the subjacent patterns that might be emphasized if the same measures would be collected for different groups. Since the library addresses sets of population, there will be situations when each group resorts to library for different informational necessities and by following different paths. As simplified example, let us suppose that a group of

population uses the references and another group uses the electronic resources; in this case, the fusion of all library users will have unsatisfactory and unusable results – half of the time is taken to resort to references and the other half is taken to resort to electronic resources. Considering these groups from different perspectives, these differences will be outlined, which will lead to much more practical results.

## **Data mining**

The goal of data mining is to explore the set of data for new and useful patterns. Data mining may have a definite goal and, therefore, we have in mind a certain goal or a certain topic area; data mining may not have a definite goal when the goal is simply to discover something interesting. Data mining includes some of the techniques that have been already discussed, as well as other instruments coming from statistics and artificial intelligence, such as neuronal networks, regression, clustering, rules and classification. Data mining represents the process of collecting a purified set of data, of generating new variables starting from the existing ones (such as the creation of a flag yes/no from a numerical variable), data sharing in sets model edification and test sets, application of techniques to sets of model edification in order to discover patterns, use of test sets to assure that the patterns are more generalizing and then conformation of these patterns by someone that knows the field [7]. These patterns will constitute then seeds for more detailed explorations that may result in new explorations, in new reports and even in new cumulative measures that may become integrant part of the measures collected on a regular basis.

In view of fulfilling data mining there are required non-cumulated data. Even if a library does not hold instruments or expertise necessary to explore own data it can assign this task to researchers if it maintains a non-cumulated version of data from their systems. By using the concept of demographic surrogate, the library will protect the identity of its users, enabling at the same time other persons to discover attached patterns from demographic data. While substituting individual users with demographic surrogates loses a part of the information, it is implied the hypothesis that the patterns associated to demographic groups may be still discovered. These patterns based on demographic information are the patterns that prove to be useful in allowing librarians to better adapt their services for groups of users. It is important to involve the library; in view of creating useful patterns the librarians have to get involved in naming the taxonomy for demographic surrogates.

A sub-area of data mining that proved to be useful in the process of digital library evaluation is the Web-based data mining. This branch of data

mining begins through a transactional log in a Web server that allows digital library services to detect the information selected by users. The challenge when investigating the exploitation through a transactional log is data purification, since the transactional logs are difficult to use. This concept was explored by Srivasta et al. [143] and was applied directly to digital libraries by Bollen et al. [8].

Data mining represents an area interwoven with bibliomining; yet, it is not directly supported by the center of data storage for bibliomining. In text mining, it is the works' context that is to be explored. The center of data storage for bibliomining, as defined herein, would not support text mining due to the fact that works are represented through demographic surrogates. Despite all these, if one wishes to carry out a project of text mining one would have to take all textual data and append fields from the center of data storage for bibliomining. This could substantially enhance the projects oriented towards text mining, as it would explore items of information comprised within works, metadata making reference to works and information on work's users.

Kostoff et al. [90] – found out that the combination between text exploitation and bibliometrics allowed a better comprehension of the community of authors than each of them used separately. They began by four works and extracted all the works they quoted. Then they resorted to text exploitation in order to search for sets of themes in the text of this corpus of papers. A key discovery consisted in the fact that the papers which quoted the original works, arising however from a different discipline, could be removed from the corpus of works that identified themselves with the adequate discipline. The disadvantage of this process consisted in the quantity of time dedicated to the manual processing required for analyzing only the four original works and quotations of first generation. Nonetheless, the work demonstrates the potential of the interweaving of concepts from bibliometrics and data exploitation for improving the comprehension of a certain situation.

## **Conceptual Framework for the Specialists in Biblioteconomy and Information Science**

Unlike the decision-making persons who need to focus upon a certain library system and would obtain the greatest benefits from a holistic understanding of that certain system, the experts in biblioteconomy aim at amplifying the understanding of biblioteconomy upon a larger scale by an activity directed towards generalisable assertions. With a view to comprehending the context for bibliomining within this research, we fix our eyes upon archaeology. Archaeology



constitutes an adequate model as the archaeologists' process of collecting artifacts in order to make assertions with respect to the humans living in a certain area is related to the process of the data warehouse creation for bibliomining. When the users wander through a digital library, they leave behind artifacts based on data of their visit. These artifacts may constitute a record for a Web log, a Proxy server login or an interrogation point used in a research within the digital library. During the phase of data warehousing, these artifacts are collected from different locations associated to the library system and are reassembled with a view to obtaining a more complete comprehension of the users who "lived" in that digital space. This concept is developed as "Internet Technology" by Nicholson [117].

In many library studies, these artifacts of the use are collected, cleaned, numbered and displayed by presentations type conference and articles, with an indication to the way in which these statistics are compared to those of other libraries. This situation is similar to archaeology until 1960: researchers focused upon a site, collected artifacts, exhibited them to others and then endeavored to operate connections between the artifacts from this and other expeditions. However, during these last 50 years, archaeologists have extended their conceptual frame by means of the "new" archaeology and of the post-processing archaeology. The new archaeology focuses upon the increase of the knowledge base, instead of simply collecting artifacts during the searches [85]. This fact may be related to the current state of digital library evaluation; rather than formulating specific questions which should advance knowledge with respect to the users of the digital library, many researchers resort to one virtual "dig" after another, collecting more and more measures with no constructive advance towards the knowledge [117]. Post-processing archeology enhanced the importance of the awareness that the situation was much more complex than the artifacts stated; issues such as the social context and the community influences had to be taken into consideration [85].

The resulting framework of the research is much more complex but it may result in creating much more generalizing knowledge assertions. Archaeologists begin by examining the artifacts available for patterns. They create generalizations from these patterns and develop hypotheses. Subsequently there are created studies for every hypothesis, and the archaeologists revisit the site with a sample methodology, collect new artifacts, talk to the people in the field and test the hypotheses. Eventually, these hypotheses are tested in other organizational frameworks for a subsequent scientific knowledge. This path is known as the hypothetical-deductive-inductive method [142].

### 1. Induction

(Recognition of the pattern by bibliomining)

### 2. Deduction

(Logical analysis of the patterns in order to obtain generalities)

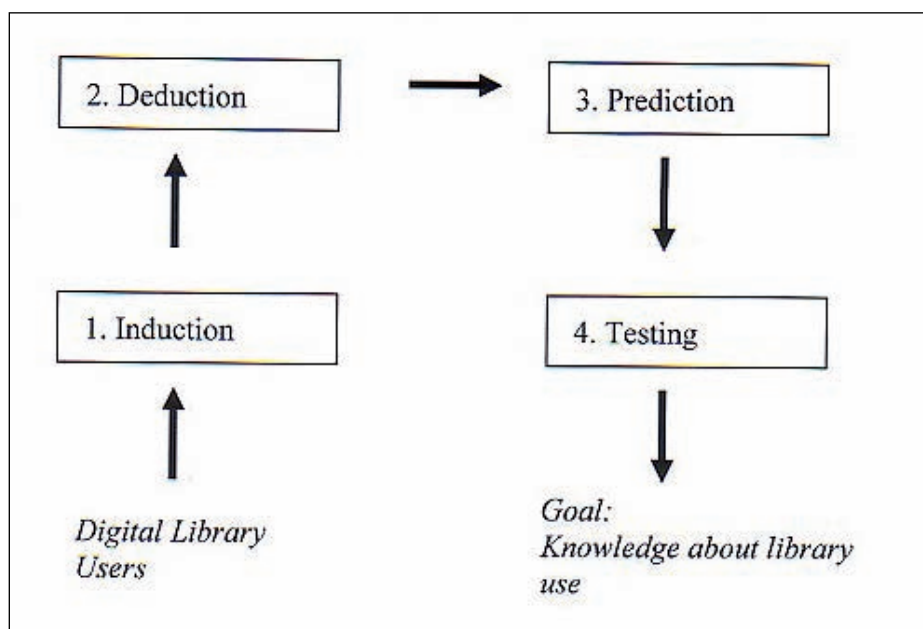
### 3. Prediction

(Creation of the hypothesis from the generalities)

### 4. Testing

(Developed research in order to test the hypotheses)

Target: Knowledge about the use of the library



**Figure 8. Framework inspired from archaeology for the research within the digital library**

Figure 8 presents an application of this framework to the research effected in the digital library and it places bibliomining in context with the rest of the research process. Similarly to the previous case, bibliomining stands for an important step within the process, nonetheless representing but a simple step in a more complex process. First of all, there are collected the artifacts of the digital library,

and bibliomining is used in order to determine the patterns. These patterns may subsequently lead to generalities, which may inspire hypotheses. Subsequently there are designated studies which should test these hypotheses; these studies may be quantitative or qualitative (or both), and they may imply addressing to the users of the digital library or collecting new data. The results emerging from these studies may lead researchers on to the path of new knowledge.

### Post-modernist Archaeology

Post-modernist archaeology avoids the process which produces generalizations about a culture and focuses instead upon the individuals of a certain culture [117]. Any culture is made up of individuals, each of them making their own decisions in life. In order to comprehend a culture, a researcher has to understand the individuals who make up that certain culture.

Due to the accent laid upon the users by the researchers in information science, such as Dervin and Nilan, important research have been advanced with respect to the importance of the users for evaluations based on artifacts during the evaluation process within libraries. Saracevic and Kantor argued the importance of the users' implication in learning the relevance and in the usefulness of the resources retrieved within a system. The holistic framework for measurements presented in Table 1 [115] presents the relation between bibliomining or the internal view of the usefulness of the system by artifacts in guise of data, and other measurement areas, which include:

- Internal view of the system, focusing upon the information and services offered;
- External view of the system, implying the users in order to find out whether the information offered meet the requirements of a certain user; and
- External view of the use, implying the users in order to find out whether the information proved useful to the user.

| Perspective                      | Subject                |  |
|----------------------------------|------------------------|--|
|                                  | System                 | Use  |
| Internal (System of the library) | Standards of procedure | Interactions registered with the interface and materials, Bibliomining |
| External (User)                  | Approximate use        | State of the knowledge<br>Value of the works                           |

**Table 3: Matrix of measurement extracted from Nicholson's measurement frame**

An advantage in examining the information consists in the implication of the users of a system in the evaluation process. We may speak of certain scenarios for the users' implication within the evaluation:

- In case of a log-on procedure, the users may comment how satisfying the results proved for their search and how useful the information was in meeting their requirements. In order to collect these viewpoints, an opinion poll effected after the interaction with the system will prevent the relevance decisions from modifying in time.
- In case of no use of a log-on procedure, the construction in the deduction methods during the information delivery process may contribute to the capture of the relevance decisions as for a piece of information, by a user. The necessary data may be captured by requesting the e-mail address and by continuing with an opinion poll about the utility of the information.
- Another approach consists in localizing the search themes for the typical users of the system. These users may operate their search according to their own needs or they may be offered searches, the measurements being subsequently collected by this process. This is the least desired scenario and it may be the only one available.

While bibliomining may contribute to the comprehension of the existent use patterns, it can't help evaluators to understand the documents deemed relevant by people (in comparison with the documents they used). Moreover, the results in bibliomining are restricted by the system resorted to; if a characteristic is not offered by a system, no information in connection with that specific characteristic can be obtained. This is the reason why bibliomining, the same way as archaeology can't offer the truth, as the truth about the search of information is in the possession of the person effecting the search. Bibliomining may retrieve information about the interactions between an individual and the system and it can retrieve nothing with respect to the user's mental state. Consequently, the collection of these measures allows the researcher to bring qualitative elements in the quantitative bibliomining process.

### **Completion of the Cycle**

The addition of post-modernist archaeology to the Hypothetical-Deductive-Inductive cycle allows researchers to place themselves behind the data in the system in order to better comprehend the users. The processes have been separated from the products in South's cycle in order to better integrate the requirements of a digital library. Making use of the entire cycle, the researchers may operate the passage from the description to the improvement of the information base about the users of the library. The significant processes of this cycle are:

- Collection: gathering the artifacts about the services of the library, about the users and resources,
- Induction: recognition of the patterns by bibliomining and visualization,,
- Deduction: logical analysis of the patterns in order to produce generalizations,
- Prediction: creation of hypotheses out of the generalities, and,
- Testing: research developed in order to test the hypotheses which imply both data and users.

It is worth knowing that we can never know the user's experience only from the artifacts within the system; not even a dialogue with user after the process offers a valid description of the user's way of thinking. The same way an archaeologist does not know the truth about the people who developed in a certain culture, we neither can know the truth about our users only looking at the artifacts. We have to work directly with the users in order to test our hypotheses with respect to their behavior.

This resulting archaeological frame is an application of the traditional scientific process. Librarians and other specialists who offer web information resources do not regularly use research projects based on hypotheses when their services are evaluated. The result is that the science of biblioteconomy and information did not follow the same development rhythm of other sciences. The advantage of the archaeological approach is that it may be of help for the non-initiated in the scientific method to conceptualize the way in which this method applies to their own information service.

### **Conclusion: Passage from Evaluation to Comprehension**

The final and longest research bibliomining can inspire consists in the improvement of the comprehension of digital libraries at a generalized, perhaps even conceptual level. These data warehouse centers will combine the resources hitherto not traditionally available for researchers under this combined form. As in the archaeological frame presented above, this is the point in which bibliomining may inspire new questions for the research process. For example, what connections may be effected between the demographic information about the user and the authors' social networks based on bibliometrics? How extensive is the influence of the written and quoted works by an institution upon the use patterns of the library services by the students? How do the use patterns differ between departments and demographic groups and what measures can the library adopt in order to personalize and improve the existing services? They may imply qualitative measures in order to support the quantitative data, or they may include the collection of other types of data from other measurement

fields which should enhance the bibliomining data. The exploration of these new questions vouched for study may lead to a thorough comprehension of this science of biblioteconomy.

Moreover, these vast warehouses of multi-system data may allow visualizations of a knowledge space by means of the comprehension of the connections among the resources resorted to or among the common points from the references transactions. Taking over the methods currently of use in the visualization of the bibliometric data, expanding them in order to include the connections among the papers and adding animation with a view to demonstrating the way in which these data modify in time may allow the understanding of the evolution of the knowledge spaces.

Due to the nature based upon artifacts of the digital library services, the decision-makers and the researchers can understand the users' information search behavior, previously available only within a restricted control environment of the research. These artifacts based upon data may be collected, deprived of identity and combined from different systems and services into data warehouses which offer thoroughness to the decision-makers and information ampleness for the users, necessary for a better understanding of the digital libraries. Resorting to bibliomining in guise of pivot of the conceptual frames allows the users and the decision-makers to cooperate for the development and solving of the research issues both of short-term interest and of long term impact.

### **Stages of a Bibliomining Project**

Bibliomining, which means data mining in the framework of a library, stands for a fragment of a more comprehensive project. The process as a whole is known as Information/Knowledge Discovery within the Data Bases (KDD). This page will enhance the way in which KDD may be made use of in the framework of the library environment.

1. Identification of the theme. The first step consists in the identification of the theme the library subject deals with. There are two types of bibliomining – predictive and descriptive. Predictive bibliomining may either predict a future event on the basis of the past and present state of affairs, or predict a current event which is difficult to visualize on the basis of a small group or of measurements carried out in the past. Descriptive bibliomining aims at describing a current situation.
2. Creation of the Data Warehouse. The data sources which may be of help in the field of the chosen theme must be identified. After this stage, the data must be collected from the adequate systems and combined within a

unique data warehouse. Moreover, the data are purified and there is carried out an operation upon the default values. This stage may occupy up to 80% of the time assigned to a bibliomining process; however, the final results greatly depend upon the accomplishment of this stage. If a librarian deems bibliomining useful, then he must dedicate a significant temporal lapse in order to create a constantly updated data warehouse, which should extract the data out of the operational system, should purify them and should deposit them upon a well fundamented basis. In the future, this investment will greatly facilitate the development and the accomplishment of the bibliomining processes.

3. Refinement of the data. Subsequently, the data adequate to the bibliomining process you focus upon are taken into consideration. New variables (ratios and classifications) may be generated from the original variables. It is possible for a detection and confrontation of the variables with the extreme values to take place. Within this stage there is necessary for an operation upon the default values ignored within the previous stages to be carried out.
4. Exploration of the data. Within this stage, the real bibliomining process unfolds. On the basis of the expected result and of the types of data, different techniques and reports are made use of in order to discover new and operable patterns (Then a miracle occurs ...).
5. Evaluation of the results. The patterns which have been discovered must be intelligible for the librarians who operate upon the field the theme affiliates itself in. If the pattern proves unusual and contravenes to the “common sense”, then it probably represents a flaw, an error with respect to the data. If this thing happens, the operator has to check the individual records with a view to discovering what generated the respective pattern. In case a predictive pattern was created, then it may be applied to a resistant sampling in order to test the reliability of the model.
6. Report and implementation. In case predictive patterns were resorted to, then they may be implemented. If necessary, they may be implemented upon a reduced sampling of real world data in order to detect their efficacy before their implementation on a large scale. There may be created and advanced reports to the members of the staff who are involved. Many a time, this may generate more questions, a fact which may send you back to step 1.

## **IV Data Mining Applied in Automated Library - a different approach –**

### **Automated analysis**

Data mining computerizes the process of thorough research of the past data with a view to discovering new information. This is the most important difference between data mining and statistics, in which case a certain model is divided by a statistician in order to treat a specific analysis problem. The same difference operates a distinction between data mining and expert systems in which case the pattern is built by an engineer out of rules extracted from an expert's experience and work.

The phases of the automated retrieval also operate a distinction from OLAP, simple questions and report tools, which are resorted to when verifying the hypotheses formulated by the user.

Data mining does not base itself upon the user in defining a specific question, only in formulating a target such as the identification of the most accessed works.

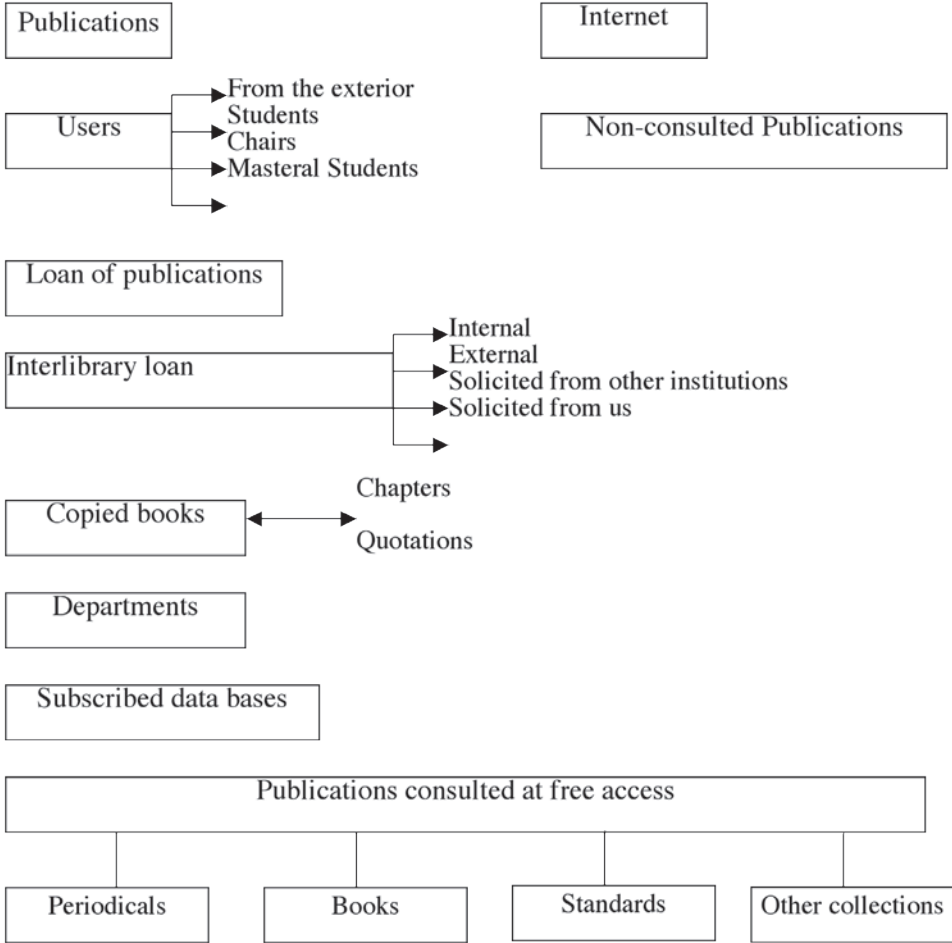
### **Large or Complex Sets of Data**

One of the attractions of data mining consists in making possible the analysis of very large sets of data within a reasonable period of time. Data mining is also convenient for complex issues which imply relatively small groups of data, in which there are however many fields or variables to be analyzed. Nonetheless, for simple and relatively minor issues of data analysis, there are simpler, cheaper or more efficient solutions.

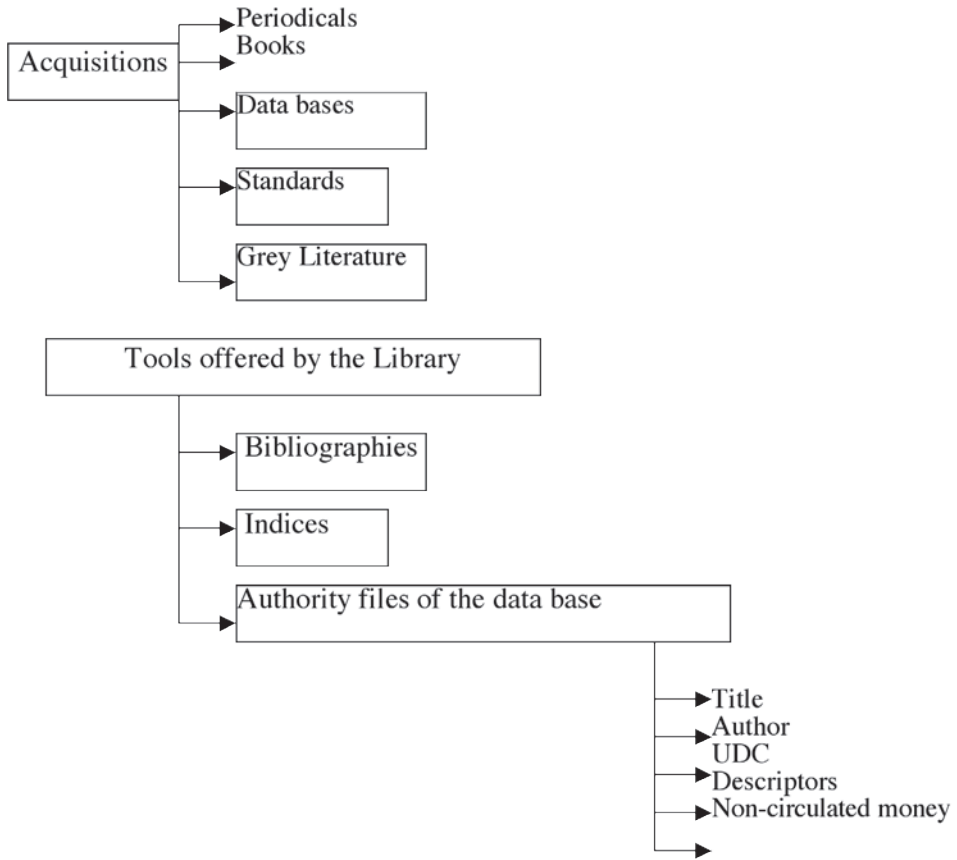


DATA WAREHOUSE

**DATA bases**



- Pattern** → Discovery of the most solicited document
- Books of the greatest interest
  - Years
  - Fields
  - Language
- Pattern:** → Non-consulted publications → Failure



**Figure 9: Data Warehouse Components**

## Discovery of Significant Patterns

The target in “data mining” consists in discovering relations among data which might provide useful meanings.

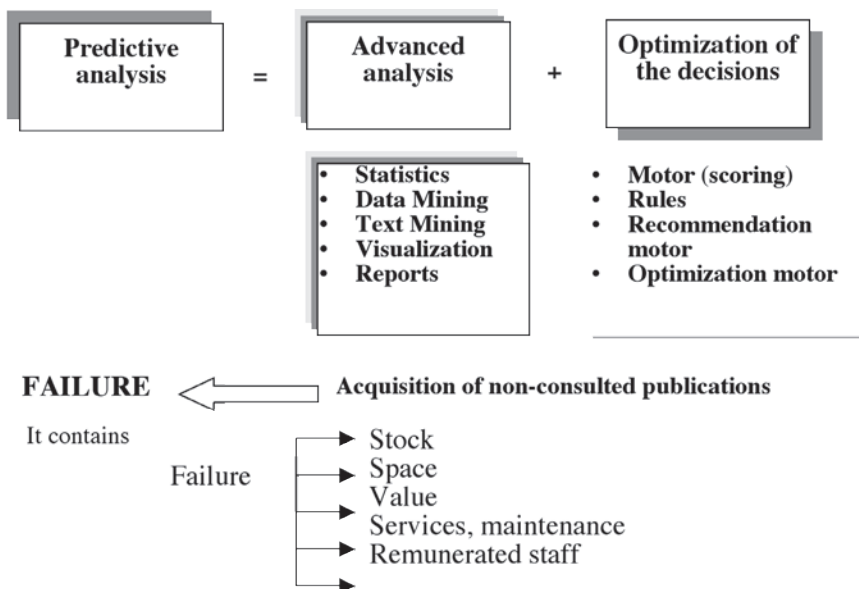
“Data mining” tools may scan data bases and may identify patterns, previously hidden, in a single step.

- One example, of pattern discovery, consists in the analysis of the most accessed works, fields, authors, publishing houses.
- Other problems of “pattern” discovery include the detection of non-circulated works with a view to avoiding the FAILURE.

The last significance of these patterns will be evaluated by an expert in the field, a marketing manager or a network administrator.

- Data Mining tools may also computerize the process of predictive information retrieval in great data bases. Questions which normally implied large analyses, may promptly find their answer in the data. A typical example of predictive problem consists in discovering the level of the targeted market sector.

Data Mining resorts to the data in the most recent editorial offers with a view to identifying the works susceptible of being consulted at a maximum level. Other predictive issues include the prevention of the failure of other forms of failure.

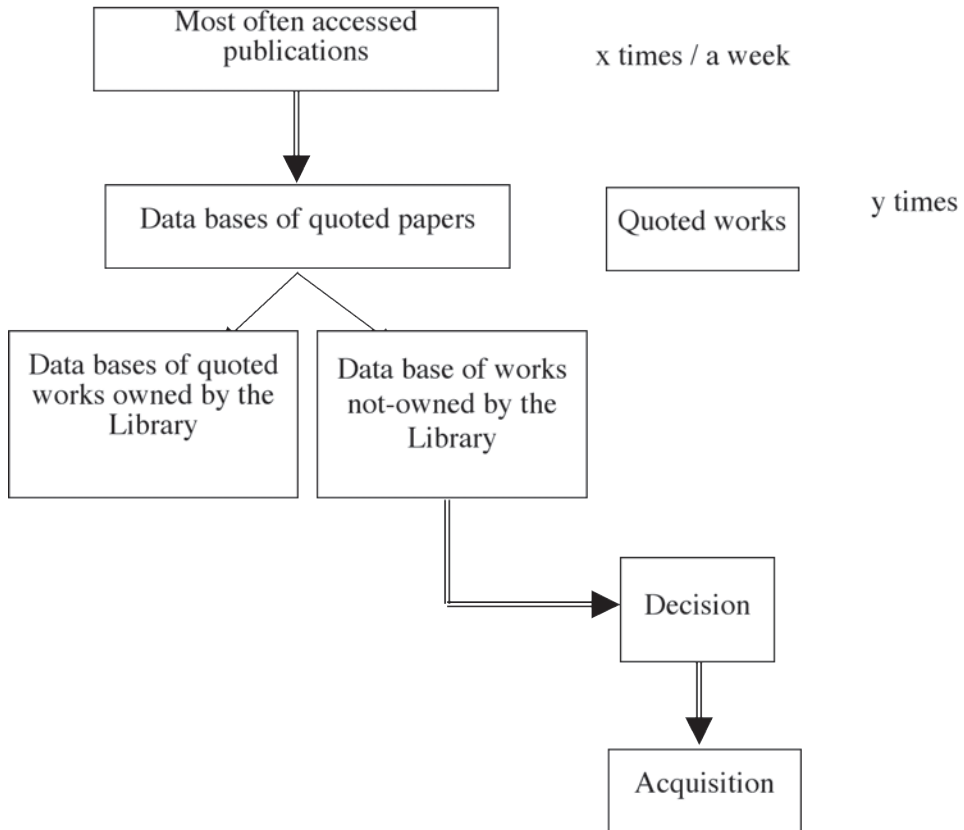


**Figure 10. Elements of predictive analysis**

A “data mining” application stands for an implementation of “data mining” technology, which solves a specific task or a research issue. Example of the application areas:

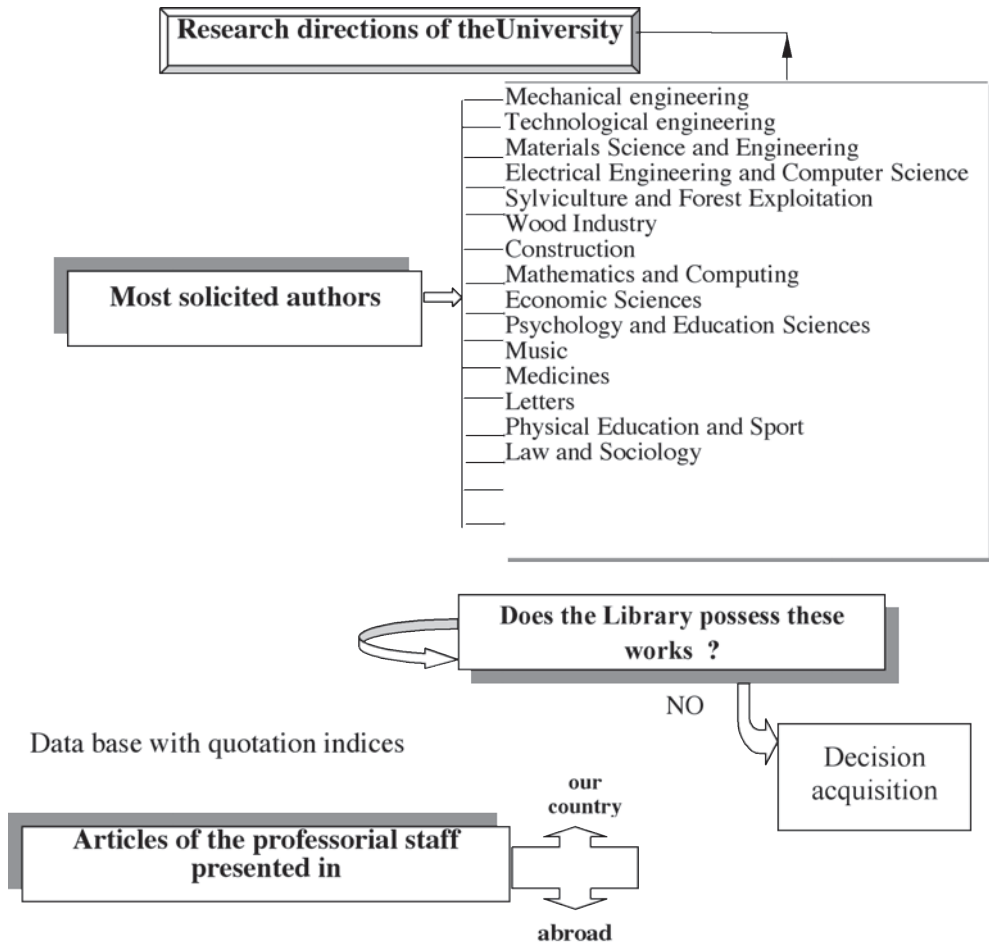
- A library may analyze the management of collections from the recent period in order to improve the objectives of meeting the users’ needs and in order to determine what publications are accessed (have had the greatest impact) during the last months. The data will include the most often quoted works and also information with respect to the works that have been quoted and do not exist within the library. The results will be distributed to the Evidence-Acquisition department, a fact that will determine the representatives of the department to revise the recommendations from the perspective of the key attributes in the process of decision. The continual analyses of the deposited data allow for the best procedures to be applied in the specific acquisition situations.

## MANAGEMENT OF COLLECTIONS



**Fig. 11 Management of collections  
Management of collection development**

- The library may control the research activity in the respective university in order to identify the fields which should manifest the greatest interest in acquiring new publications. By resorting to data mining, there may be easily identified the most solicited authors in different fields, the professorial staff's articles appeared within our country and abroad, there may be created a data base with quotation indices.



**Figure 12: Management of collections. Development and directions of research**

Data mining techniques may give up the benefits of the existing software and hardware platforms automation in order to enhance the value of the existing informational resources, and they may be implemented upon new products and systems once they are available on-line. When these techniques are implemented upon customer/server processing systems or parallels of high performance, they may analyze massive data bases and may answer questions such as: What department, field proves itself most satisfied with the offer of the library?

Data Mining technology is prepared for applications, as it is supported by three technologies that are currently sufficiently developed, which are:

1. Massive data collections
2. Powerful multi-processor computers
3. Data search algorithms

The main components of data mining technology have been under development for decades, in research fields such as statistics, artificial intelligence and learning of intelligent machines. Currently, the development of these technologies in compliance with relational data bases browsers of high performance and the ample efforts of data integration make these ones extremely practical for the fields of data depositing.

The key for understanding the different aspects of data mining consists in the distinction among the applications, operations, techniques and algorithms of data mining.

|                     |   |
|---------------------|---|
| <b>Applications</b> | Data bases market<br>Classification of the users  |
| <b>Operations</b>   | Classification and prediction<br>Clustering<br>Analyses of associations<br>Prevision                    |
| <b>Techniques</b>   | Neuronal networks<br>Decision trees<br>Close neighbour algorithms<br>Naïve Bayesian<br>Cluster analysis |

**Table 3: Data Mining. Elements**

## **Operations**

An application which resorts to data mining technology will implement one or more data mining operations (many times called tasks). Every operation reflects a different modality for the distinction of the patterns or of the orientations within a complex data set.

## **Classification and Prediction**

Classification stands for the operation which is most used by the data mining commercial instruments. It is an operation that helps organizations discover

certain patterns in vast and complex data banks with a view to solve specifically business issues.

**An example of classification:** University Library wants to identify those users who might be interested in a new opportunity of data bases acquisition. It also launched formerly on the market such a product and it possesses data that indicate which of its users answered the previous offer. The goal consists in understanding what factors identify the users who positively answered the offer, so that the effort made in marketing and in acquisition should be focused more efficiently.

There is, in the users' registration, a field that it set as being "true" or "false" according to the users' answer to the offer. The field is denominated "target field" or "dependent variable" for classification. The goal is to analyze in what way other attributes of the users (such as the field they study, the year, the school performances, the number of the visits to the library) influence the class they belong to (the class indicated by the target field). This piece of information will be deposited in other fields from the user's registration. The distinct fields included within the analysis are identified as fields or independent or predictive variables.

**Classification techniques:** The manner in which "data mining" instruments analyze the data, and the type of information they offer, depends upon the techniques resorted to. The most common techniques of classification are: the decision trees and the neural networks. If a decision tree is made use of, we shall have at our disposal a set of ramified conditions which successfully divide users into groups, defined by the values from the independent variables. The aim is being able to produce a set of rules or a certain type of model.

By contrast, a neuronal network identifies the class a certain user belongs to, not being able however to state why. The factors which determine the classification are not available for analysis, remaining however involved within the network. Another set of techniques resorted to for classification are the k-nearest neighbor algorithms.

## Softwares for Data Mining classification

1. Multiple approaches, typically including both a decision-tree and a neural network models, as well as some way to combine and compare them.
2. Classification with Neural networks
3. Classification with Decision tree or Rule-based approaches
4. Bayesian and Dependency Networks

5. SVM (Support Vector Machines)
6. Other approaches, including Rough sets and Genetic algorithms.
7. Analysis of classification results, ROC curves, and more

## 1. Software for Classification using multiple approaches

### **commercial:** | free

- Affinium Model Suite, includes linear regression, logistic regression, CHAID, neural networks, and genetic algorithms.
- Clementine from SPSS, leading visual rapid modeling environment for data mining. Now includes Clementine Server.
- KINOSuite PR, extracts rules from trained neural networks.
- Knowledge Studio, featuring multiple data mining models in a visual, easy-to-use interface.
- MarketMiner automatically selects the best mining technique using: Statistical Networks, Logistic and linear regression, K nearest neighbors, and Decision trees (C4.5).
- MLF: machine learning framework for Mathematica, the multi-method system for creating understandable computational models from data.
- Oracle 9i Data Mining, embeds data-mining functionality into the Oracle9i database, for making classifications, predictions, and associations.
- Polyanalyst, features multiple classification algorithms: Decision Trees, Fuzzy Logic, and Memory Based reasoning.
- Predictive Dynamix Data Mining Suite integrates statistical, graphical, and ROC analysis with neural network, clustering, and fuzzy models.
- PredictionWorks, includes decision tree (gini, entropy, C4.5), logistic regression, k nearest-neighbor, naive bayes and linear regression. Free test over the web!
- Previa Classpad, provides an interactive environment for classification using neural networks, decision trees, and bayesian networks.
- prudsys DISCOVERER: non-linear decision trees (NDTs) and sparse grid methods for classification
- Purple Insight MineSet, offering several classification methods.
- STATISTICA Data Miner

### *free:*

- JAM, Java Agents for Meta Learning (applications to fraud and intrusion detection).
- MLC++, a library of C++ classes for supervised machine learning, including multiple classification algorithms.
- SIPINA-W, produces decision graphs and trees. Includes several classification methods. (Win). Shareware



## 2. Neural Network Software for Classification

*sites:*

- Neural Network FAQ list, includes free and commercial software, maintained by Warren Sarle of SAS.
- Portal for Forecasting with neural networks, including software, data, and more.

*commercial | free*

- Alyuda NeuroIntelligence, supports all stages of neural network design and application.
- BioComp iModel(tm), self-optimizing, non-linear predictive model development.
- COGNOS 4Thought, predictive modeling tool offering Effectiveness measurement, What-if analysis, and Forecasting
- BrainMaker, a fast neural network system, now with MMX acceleration (Win 95/NT)
- KINOsuite
- MATLAB Neural Net Toolbox, a comprehensive environment for neural network research, design, and simulation within MATLAB
- NeuroSolutions, powerful and flexible neural network modeling software.
- NeuroXL, neural networks software for classification and prediction of simple and complex data in Excel.
- NeuralWorks Predict Predict 3.0 and Professional II/PLUS.
- SPSS Neural Connection 2, with Bayesian Network, Data Output Tool, model weights and more.
- STATISTICA Neural networks, comprehensive and user-friendly nn application with many charting options, network architectures and training algorithms.

*free and shareware:*

- NuClass7, freeware for fast development, validation, and application of neural and conventional classifiers including multilayer perceptron, functional link net, piecewise linear net, nearest neighbor classifier, self organizing map.
- Sciengy RPF(tm) for Data Mining, a free experimental Windows application self-organizing neural networks, a convenient user interface and ability to work with text data files.
- Tiberius, MLP Neural Network for classification and regression problems.

### 3. Decision Tree Software for Classification

*commercial* | free

- AC2, provides graphical tools for data preparation and building decision trees.
- Alice d'Isoft 6.0, a streamlined version of ISoft's decision-tree-based AC2 data-mining product, is designed for mainstream business users.
- C5.0/See5, constructs classifiers in the form of decision trees and rulesets. Includes latest innovations such as boosting.
- CART 5.0 decision-tree software, multiple winners of KDD Cup. Advanced facilities for data mining, data pre-processing and predictive modeling including bagging and arcing.
- Compumine Rule Discovery System, has a complete coverage of rule-based predictive modeling methods for both classification and regression, with a user-friendly interface.
- DrawBTree, Binary Decision Tree flowcharting software, as well as examples of how these flowcharts are used in documents.
- DTREG, generates classification and regression decision trees; finds optimal tree size; supports variable costs, priors and variable weights. Download free demo version.
- Decisionhouse, provides data extraction, management, pre-processing and visualization, plus customer profiling, segmentation and geographical display.
- Fair, Isaac Model Builder for Decision Trees, advanced tree-building software that leverages your data and your business expertise to guide you in strategy development.
- KnowledgeSEEKER, high performance interactive decision tree analytical tool.
- Neuscience's aXi.DecisionTree, ActiveX Control for building a decision tree. Handles discrete and continuous problems and can extract rules from the tree.
- PolyAnalyst, includes an information Gain decision tree among its 11 algorithms.
- Purple Insight MineSet, offering decision trees and other classification methods.
- Shih Tree Builder, Classifier/regression/probability tree, manual/automatic split, pruning, priors, misclassification costs, train/test parallel monitoring, mixed algorithms, exhaustive search of possible splits.
- SPSS AnswerTree, easy to use package with CHAID and other decision tree algorithms. Includes decision tree export in XML format.
- XpertRule Miner (Attar Software), provides graphical decision trees with the ability to embed as ActiveX components.

*free and shareware:*

- C4.5, the “classic” decision-tree tool, developed by J. R. Quinlan, (restricted distribution).
- BTreePseudo, pseudo-code for a decision-tree program.
- Classification Tree in Excel, from Angshuman Saha
- DM-II system, includes CBA for classification based on associations, and many more features.
- GAtree, genetic induction and visualization of decision trees (free and commercial versions available).
- IND, provides Gini and C4.5 style decision trees and more. Publicly available from NASA but with export restrictions.
- LMDT, builds Linear Machine Decision Trees (based on Brodley and Utgoff papers).
- Mangrove, a tool for visualization of decision tree learning,
- OC1, decision tree system continuous feature values; builds decision trees with linear combinations of attributes at each internal node; these trees then partition the space of examples with both oblique and axis-parallel hyperplanes.
- ODBC MINE, analyzes ODBC data bases using C4.5, and outputs simple IF. ELSE decision rules in ascii.
- PC4.5, a parallel version of C4.5 built with Persistent Linda (PLinda) system.
- SMILES, advanced decision tree learner, with new splitting criteria, non-greedy search, extraction of different solutions, boosting, cost-sensitive learning, and more.
- Random forests from Leo Breiman, a combination of tree predictors such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest.

### **Rule-based approach**

*commercial:*

- Datamite, enables rules and knowledge to be discovered in ODBC-compliant relational data bases.
- PolyAnalyst, supports decision tree, fuzzy logic rules, and other classification algorithms.
- SuperQuery, business Intelligence tool; works with Microsoft Access and Excel and many other data bases.
- WizWhy, automatically finds all the if-then rules in the data and uses them to summarize the data, identify exceptions, and generate predictions for new cases.
- XpertRule Miner (Attar Software) provides association rule discovery from any ODBC data source.

*free:*

- CBA, mines association rules and builds accurate classifiers using a subset of association rules.
- CN2, inductively learns a set of propositional if...then... rules from a set of training examples by performing a general-to-specific beam search through rule-space.
- KINOsuite-PR extracts rules from trained neural networks.
- PNC2 Rule Induction System, Windows software tool that induces rules using the PNC2 cluster algorithm.

## 4. Bayesian and Dependency Networks Software

**Overview pages** | commercial | free

- Kevin Murphy's Bayesian Network Software Packages page
- Google's list of Bayes net software.

**commercial:**

- AgenaRisk, visual tool, combining Bayesian networks and statistical simulation (Free one month evaluation).
- Analytica, influence diagram-based, visual environment for creating and analyzing probabilistic models (Win/Mac).
- AT-Sigma Data Chopper, for analysis of data bases and finding causal relationships.
- BayesiaLab, complete set of Bayesian network tools, including supervised and unsupervised learning, and analysis toolbox.
- Bayesware Discoverer 1.0, an automated modeling tool able to extract a Bayesian network from data by searching for the most probable model
- BNet, includes BNet.Builder for rapidly creating Belief Networks, entering information, and getting results and BNet.EngineKit for incorporating Belief Network Technology in your applications.
- DXpress, Windows based tool for building and compiling Bayes Networks.
- Ergo(tm), Bayesian Network Editor and Solver (Win and Mac demos available)
- Flint, combines bayesian networks, certainty factors and fuzzy logic within a logic programming rules-based environment.
- HUGIN, full suite of Bayesian Network reasoning tools
- KnowledgeMiner , uses self-organizing neural networks to discover problem structure (Mac platform)
- Netica, bayesian network tools (Win 95/NT), demo available.
- PrecisionTree, an add-in for Microsoft Excel for building decision trees and influence diagrams directly in the spreadsheet

**free:**

- BAYDA 1.0
- Bayesian belief network software (Win95/98/NT/2000), from J. Cheng, including
  - BN PowerConstructor: An efficient system for learning BN structures and parameters from data. Constantly updated since 1997.
  - BN PowerPredictor: A data mining program for data modeling/classification/prediction. It extends BN PowerConstructor to BN based classifier learning.
- Bayesian Logistic Regression Software, for large-scale Bayesian logistic regression (Windows and Linux)
- Bayesian Network tools in Java (BNJ): an open-source suite of Java tools for probabilistic learning and reasoning (Kansas State University KDD Lab)
- FDEP, induces functional dependencies from a given input relation. (GNU C).
- GeNIe, decision modeling environment implementing influence diagrams and Bayesian networks (Windows). Has over 2000 users.
- JavaBayes
- jBNC, a Java toolkit for training, testing, and applying Bayesian Network Classifiers.
- MSBN: Microsoft Belief Network Tools, tools for creation, assessment and evaluation of Bayesian belief networks. Free for non-commercial research users.
- PNL, Intel Open-Source Probabilistic Network Library
- Pulcinella, tool for Propagating Uncertainty through Local Computations based on the Shenoy and Shafer framework. (Common Lisp)

## **5. SVM (Support Vector Machines) Software for Classification**

*commercial:*

- EQUBITS Foresight(tm), SVM-based Predictive Modeling.
- KXEN, Components, based on Vapnik's work on SVM.

*free:*

- BSVM, a decomposition method for support vector machines (SVM) for large classification problems.
- Gist, web interface to SVM
- Kernel Machines and related methods website
- LS-SVMlab, Least Squares - Support Vector Machines Matlab/C Toolbox
- LIBSVM, a support vector machines (SVM) library for classification
- pcSVM, a framework for support vector machines, including a pcSVM demo program which trains SVM on generated 2-D classification problems and draws the computed decision line together with the identified support vectors.

- SVM-light, popular implementation of Support Vector Machines (SVMs) by Thorsten Joachims.
- SvmFu 3, SVM package in C++, under GPL (source only, requires compilation)

## 6. Genetic Programming, Rough Sets and other classification software

*commercial:*

- Datalogic, professional tool for knowledge acquisition, classification, predictive modelling based on rough sets.
- Evolver, genetic programming
- GAtree, genetic induction and visualization of decision trees (free and commercial versions available).
- Genalytics GA3, uses genetic programming to dynamically build predictive models.
- gepsoft APS 3.0, an extremely flexible modeling tool for Function Finding, Classification, and Time Series Prediction based on Gene Expression Programming; generates Visual Basic; C++ and more.
- MARS, J. Friedman's automated logistic regression for binary classification problems. Automatic missing value handling, interaction detection, variable transformation.
- WINROSA, automatically generates fuzzy rules, based on the fuzzy ROSA method (Rule Oriented Statistical Analysis).

*free:*

- Grobian, user-friendly software to analyse data with rough set technology (C++).
- Rough Set Exploration System (RSES), contains classification based on rough sets, decision tree, LTF-network, instance based classification and data discretization (free for non-commercial use).
- TiMBL 2.0, nearest neighbour approach

## 7. Software for Analysis of Classification

- Analyse-it, Method evaluation & validation software for Excel. Includes ROC curves, method comparison, NCCLS imprecision, reference ranges, and NCCLS linearity. Free 30-day trial.
- MedCalc, Windows statistical program for biomedical researchers; includes ROC analysis and more.
- ROCON, a tool to aid ROC analysis of machine learning classifiers.

***Comprehension and Prediction:*** Sophisticated classification techniques support us in retrieving new patterns in large and complex sets of data. Therefore, classification stands for a powerful help in comprehending a certain problem, even if it represents the installment of the answers to a promotion of the library services.

Under certain circumstances, it provides an improved comprehension. This may suggest new initiatives and may provide information which improves the process of decision-making in the future. However, in many cases, the reason for the development of a precise classification model consists in improving the prediction ability.

They say a classification model should be accomplished in compliance with historic data, in which case the result for every registration is known. This is subsequently applied to a new set of non-registered data, with a view to predict the result for every registration.

There are significant differences between classifying data with a view to understanding the existing users' behavior and resorting to the respective classification in order to predict the future behavior.

As for sets of data from the past, it is often possible to produce a set of rules or a mathematical function which should accurately describe every registration.

Building a good predictive model implies avoiding the over-charge by testing and harmonizing the model, in order to ensure the fact it may be generalized to the new data.

## **Clustering**

Clustering stands for an operation that cannot be controlled. It is resorted to where we want to retrieve groups of similar records in the data we have at our disposal, with no other precondition the respective similarity implies. Clustering is used in order to identify the groups of interest from a data base of the users, groups which have not been previously recognized. For example, it may be used in order to identify similarities in the users' information needs, with a view to inventing and selling new services of documentary information.

Clustering is obtained, as a rule, by means of statistic methods, such as the k-means algorithm, or a special form of the neuronal network called "Kohonen network map". Whatever should be the method resorted to, the basic operation remains the same. Every registration is compared to a set of existing clusters, which are defined by their "center". A certain registration is due to the cluster that it is nearer to, and the former changes the value which defines the cluster. Several steps are made until a set of data has re-allotted themselves the records and has modified the centers of their clusters until the discovery of the optimal solution.

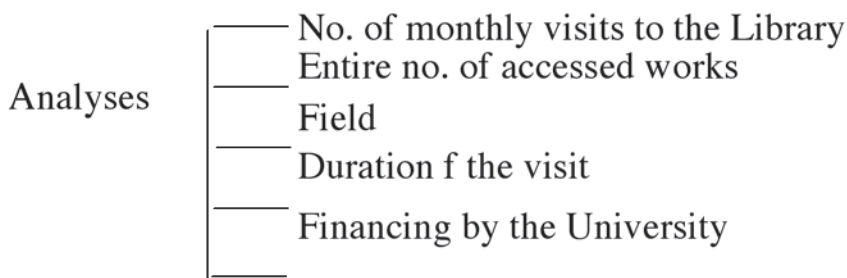
For example, searching for clusters among the library users, there might be needed to analyze more factors, among whom the number of monthly visits to the library, the whole number of the accessed works during one visit, the time of the visit and the financing by the university.

Clustering is often interpreted as an exploration exercise before continuing to search the data by means of the classification technique. For this reason, a good visual knowledge represents a further support for clustering: it makes you deepen into the work with the clusters, with a view to recognizing if the identified clusters make sense and if they can be of help in the context of the business.

## Classification technique

### Clustering

|                      |
|----------------------|
| <b>Library Users</b> |
|----------------------|



### Analysis of the Association and Sequential Analysis

The analysis of the association stands for an uncontrolled form of “data mining” which searches for connections among the registrations from a set of data. The analysis of the association is sometimes defined as “analysis of the accessed sites, which is its most used application. The aim is to discover, for example, what field is mostly accessed.

### Prognosis

Prognosis is often undertaken with regression functions – statistic methods for the examination of the relations among the variables in order to predict the next value. The statistic packages, such as SAS and SPSS, provide an extensive range of such functions which may manipulate more and more complex issues.



Nonetheless, such statistic functions usually imply a significant knowledge of the techniques resorted to and of the original conditions applied to their implementation. Data mining tools may also provide functions for prognosis. In particular, the neuronal network was used upon vast areas for the prognosis of the stocks on the market.

An important distinction may be operated between two different types of the prognosis issue.

The simpler problem consists in the prognosis of a single continuous value on the basis of series of unordered examples.

## Techniques

An operation of the type data mining may be obtained by resorting to a number of techniques or methods.

Every technique may be itself implemented in different manners, by resorting to different algorithms.

## Clustering Algorithms

The analysis of the clusters stands for a process of identification of the clusters existing between articles on the basis of the resemblance and distinctions among them. Unlike classification, clustering does not imply the previous identification of a target variable. An algorithm checks the potential groupings from the multitude of data and endeavors to obtain an optimal delimitation of the articles on the basis of those groupings.

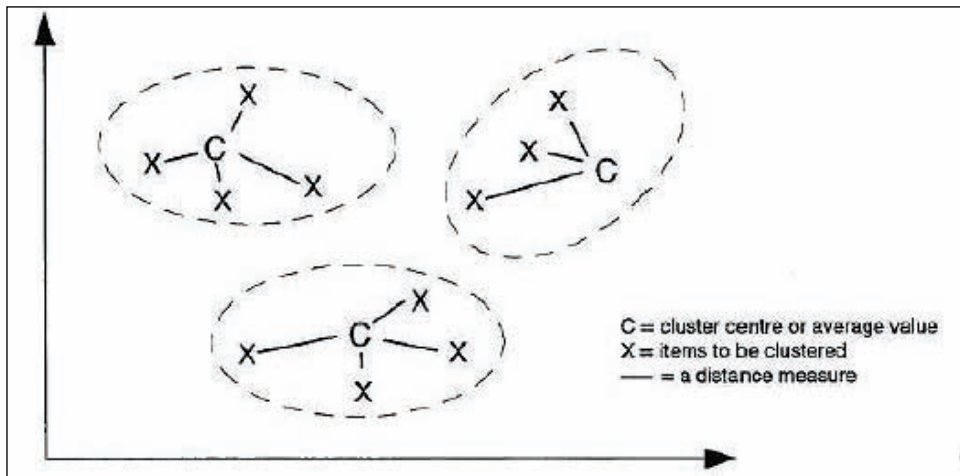


Figure 14: Grouping of the Clusters

Clusters are usually placed around a “center” or average value. The original manner of defining and regulating the centers varies according to the algorithms. One manner consists in beginning with a random set of centers, which are subsequently regulated, added and eliminated while the analysis advances.

In order to identify the articles from a cluster, we must resort to a sort of measure which equates the resemblance among the articles in a cluster and the distinctions from the articles in other clusters. The resemblance and the distinctions among the articles are usually measured as the distance between them and the others and to the centers of the clusters from the multidimensional space, where every dimension stands for one of the variables which undergo the comparison.

## **Nearest Neighbour**

The nearest neighbor (more exact k-the nearest neighbor, as well as k-NN) stands for a prediction technique adequate for the classification models.

Unlike other predictive algorithms, the stimulation data are not scanned or processed for the creation of a model. In exchange, the stimulation data stand for the model. When a new case is presented, the algorithm operates a search in all the data, in order to find a subset of cases which mostly resemble one another, and uses it in order to predict the consequence.

There are two main drivers in the algorithm k-NN: the number of the nearest cases which are to be used (k) and a metrics in order to measure what really means the nearest.

Every use of the algorithm k-NN implies what we call entire positive value of k. This determines how many cases already existent have been already studied when a new case is predicted. K-NN refers to a family of algorithms that can be understood as 1-NN, 2-NN, 3-NN and so on and so forth. For example 4-NN indicates that the algorithm will make use of the nearest 4 cases in order to predict a new case.

k-NN is based on the concept of distance and this implies a metrics in order to determine the distances. All metrics have to result in a specific number as for the comparisons. Any metrics resorted to is arbitrary and extremely important. It is arbitrary because it exists no control definition of what is called a “good” metrics. It is important because the choice of a metrics has a strong impact upon the predictions. This means an expert in business is needed in order to contribute to the determination of a good metrics. In order to classify a new case, the algorithm calculates the distance from the new case to each case (line) of the stimulation data. It is predicted for the new case to have the same consequence as the prevailing consequence in the k close cases from the stimulation data.

## **Neuronal Networks**

A key difference between the neuronal networks and any other technique is that neuronal networks operate exclusively on the numbers. It results that any non-numeric data in independent columns must be converted into numbers before being able to use the data in a neuronal network.

## **Naïve-Bayes**

Naïve-Bayes constitutes a classification technique that is both predictive and descriptive. It analyzes the relation between every independent variable and every dependent variable in order to derive a conditional probability for every relation.

Naïve-Bayes implies a single examination of the learning set in order to generate a model of classification, which makes it the most efficient data mining technique. However, Naïve-Bayes does not operate with continuous data, so as any variable, independent or dependent, which possesses continuous values, has to be encapsulated.

## **Decision Trees**

Decision trees constitute one of the most used data mining techniques. They are easy to use, the results are intelligible by a regular user, they address a wide range of classification issues and they manifest efficacy in processing o large volume of data.

An algorithm based on decision trees divides the set of data with a view to building a model that classifies every registration in terms of target field or variable. An example consists in a decision tree which divides a set of data according tot the way the users did or did not access a certain site.

The most used algorithms on the basis of decision trees are CHAID, CART and C4.5. CHAID (Chi-square automatic interaction detection) and CART (Classification and Regression Trees) were developed by statistics. CHAID may produce an tree with multiple sub-knots for every division. CART implies less preparing of the data than CHAID, but it always divides the set in only two parts. C4.5 comes from the universe of the intelligent machines capable of learning and it is based upon the theory of information.

In order to generate a decision tree out of a set of stimulation data, it is necessary to progressively divide the data in smaller sub-sets. Every iteration takes into consideration the data from a single knot. The first iteration processes the root-knot, which contains all the data. The iterations which follow work upon derived knots which will contain sub-sets of data. In every iteration we must choose that independent variable which divides the data with most efficacy.

This means the produced sub-sets must be as “homogenous” as possible towards the dependent variable.

## **V Text Mining Applied in Digital Library**

The information in electronic format has become omnipresent. The Internet provides its users with large amounts of information in any field of interest. Within this context the apparition and development of systems assisting and facilitating the access to information has become compulsory.

These systems may be of real help with regard to the automated classification of documents, information, information extraction, summarizing, browsing efficiency, increase of machines understanding, etc.

The greatest part of these systems is framed in the field of Text Data Mining. The information extraction is one of them, having a great importance in the nowadays context.

### **Web application**

Through this project we aimed at conceiving a programme of knowledge extraction in HTML documents. The program recognizes events of a certain type (weather, sport, politics, text data mining, etc) taking into account the way it will be driven (the concept-based dictionary it chooses). These events may be provided to the user or the whole context where the event occurred can be extracted in order to indicate the initial form where the event was framed.

This project aims to be of great help in case the information is searched in more web pages (Yahoo, Google, CNN site, etc.) for a certain area and a certain period when the project can be useful. For each type of events (conferences, specialized education, higher education) there is a concept-based dictionary. Conceiving such a dictionary is a very demanding task. We are endeavoring to think up an extensible system to allow the extraction of specific events. At present the automated built dictionaries are of greatest interest, through learning algorithms.

## Definitions and acronyms

Text Data Mining = TDM

Concept-Based Dictionary = DC

Part of Speech Tagger = POST

Text Data Mining = Field of extracting new items of knowledge from the text

Concept = assembly of all items of information of a certain type that follows a certain syntax imposed by concept (ex: {**information science**, university, sem. I}, {**information science**, university RO, sem.II}, {**information science**, X, Y, for any X and Y valid, plus a series of valid syntaxes valid for the concept of information science and attributes X and Y)

Concept-based dictionary = Assembly of concepts for a particular field

POST = Applications having as aim the annotation of a text with specific speech parts specific to each word in the sentence

WordNet = A data base with English language words, with classification at semantic level

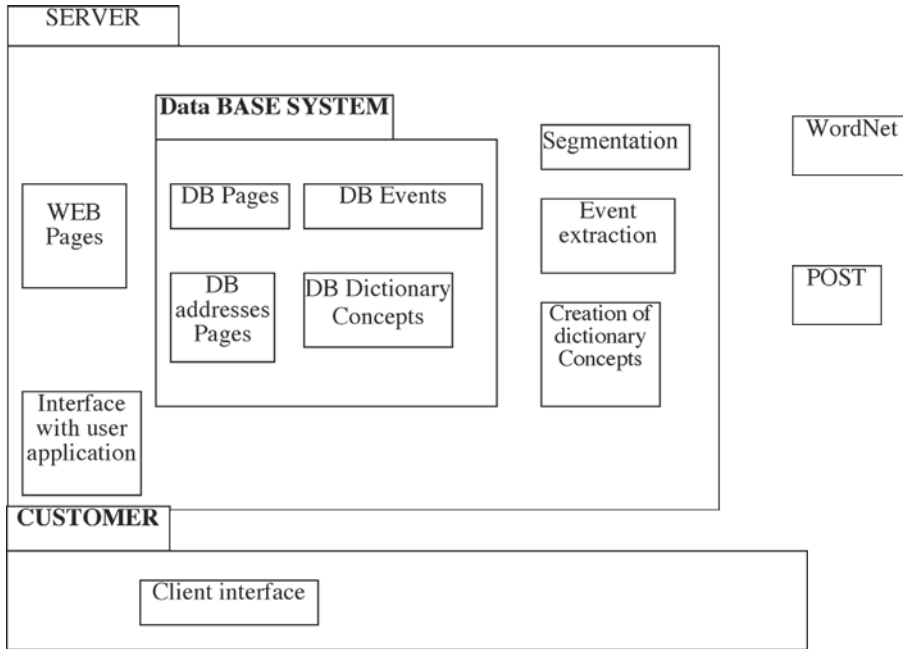
Event = an instance of a concept, that is an assembly of information that verifies – respects the structure imposed by a concept (ex: {**information science**, university, sem.I})

Segment = part of a text that represents a sentence or a phrase

## General outlines

The field of events extraction from text with natural language (for example Web pages) is in continuous development. The researchers' meetings in the field were of great interest, one of the most important one being MUCs (Message Understanding Conference). The systems developed and tested within these conferences were oriented towards the discovery of terrorist messages, of attacks upon messages on the Internet. Taking into account the fact that the text is in natural language, which has an increased complexity, and, therefore, a formalization is almost impossible, the results obtained in the correct identification of events is quite good, raising up to 80-90% in some cases.

## Description upon component parts Diagram moduli

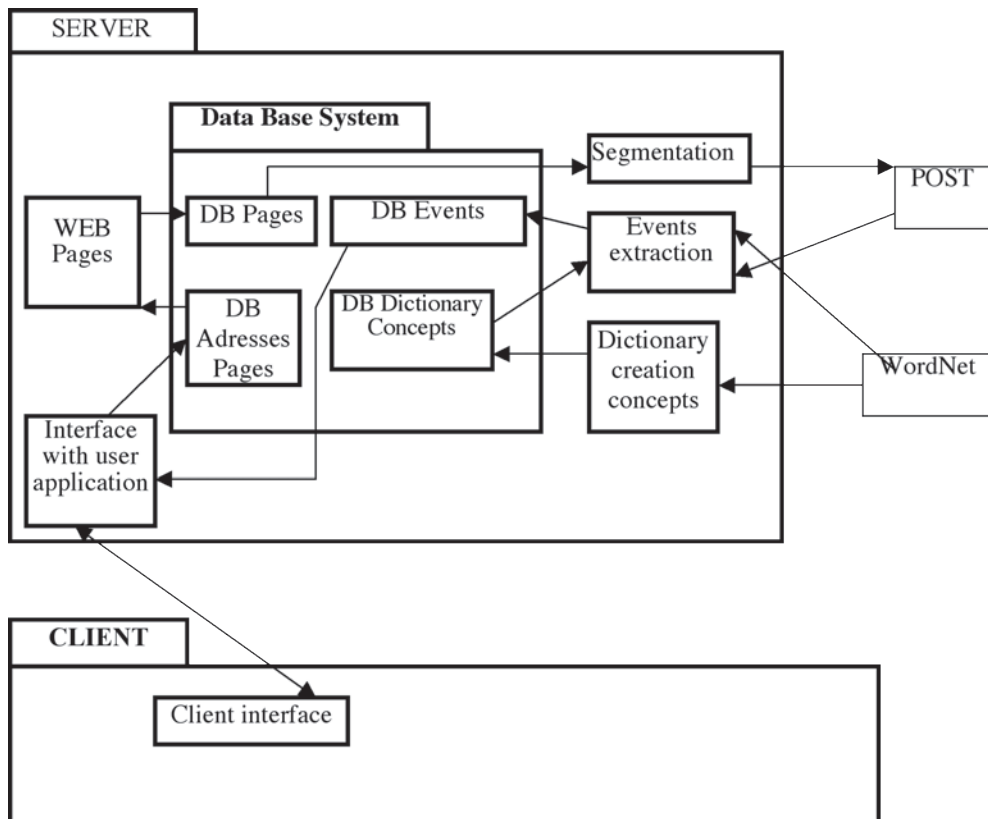


**Figure 15: Components moduli diagram**

## Modules and their components

| MODULE                    | FUNCTION  |
|---------------------------|---|
| Pages Addresses Data Base | It keeps the list of Web pages for a certain field on which the events are searched   |
| Pages Data Base           | It keeps the Web pages, indexed and on fields categories  |
| Events Data Base          | It keeps the identified events, indexed according to fields, time locations and other criteria  |
| Concept-based Data Base   | It keeps the concepts that will be searched and indexed on fields   |
| Web pages fetching        | It fetches Web pages from Pages Addresses Data Base and it stores them in the Pages Data Base. This module may recursively fetch the pages indicated by links in the current page, from the same server, till a certain fetching depths |
| Segmentation              | It divides the text from the Web pages within the Pages Data Base in segments   |
| POST                      | This module receives the segments and annotates them with the speech parts corresponding to words in segments   |
| Events extraction         | It extracts events from annotated segments, received from the POST module, on the basis of concepts in the Concept-based Dictionary and of WordNet, adding them to the Events Data Base   |
| Concept-based Dictionary  | It creates the concepts specific to a certain field, through learning algorithms, then it adds them to the Concept-based Data Base  |
| WordNet                   | It provides relationships between words of semantic nature  |
| User interface            | It receives requests for events searching, requests received from the Events Data Base  |
| Customer interface        | It assures the user interface, the introduction of events queries, options setting, results display according to various criteria   |

**Table 4: Component moduli**



**Figure 16: Moduli diagram**



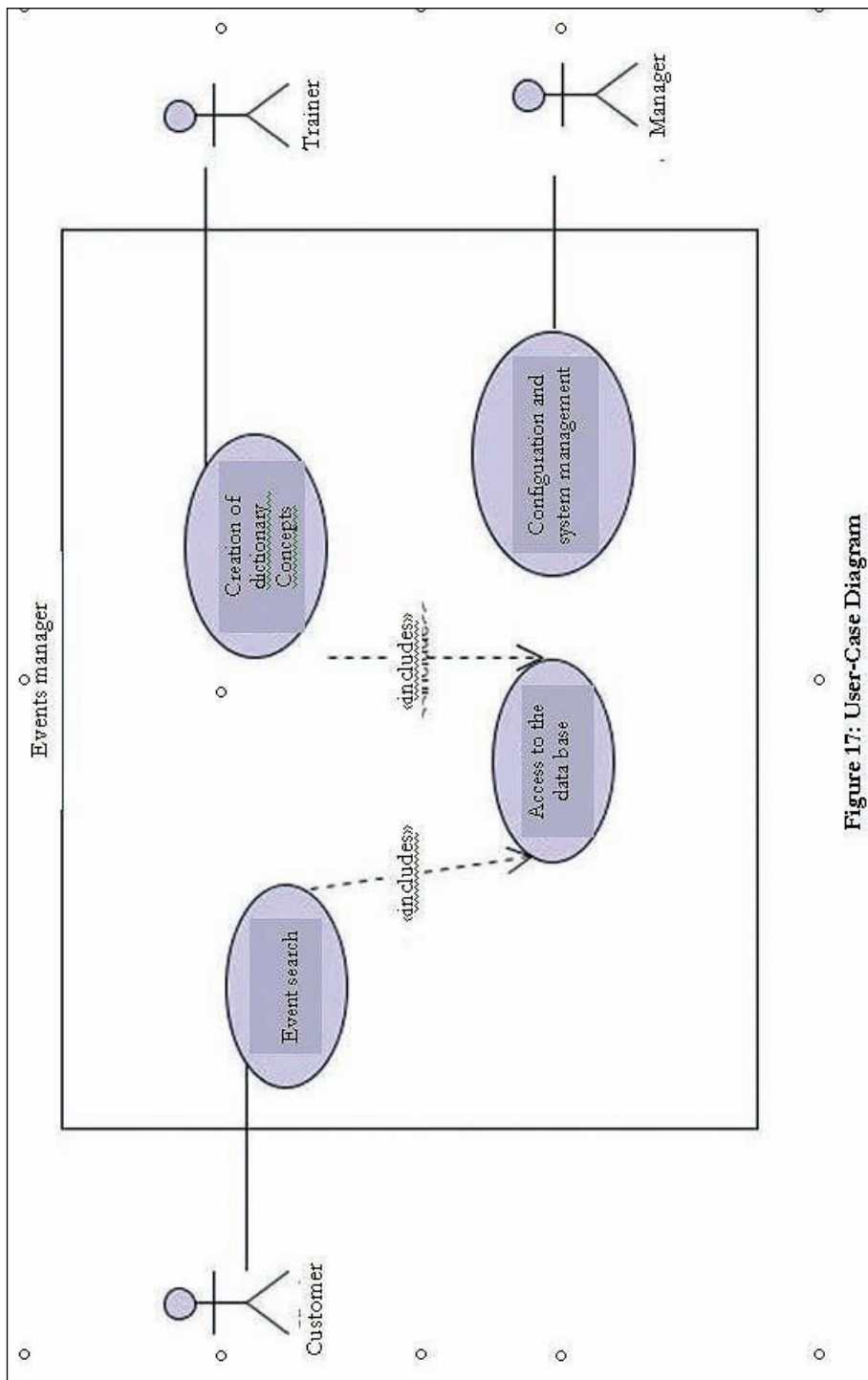


Figure 17: User-Case Diagram

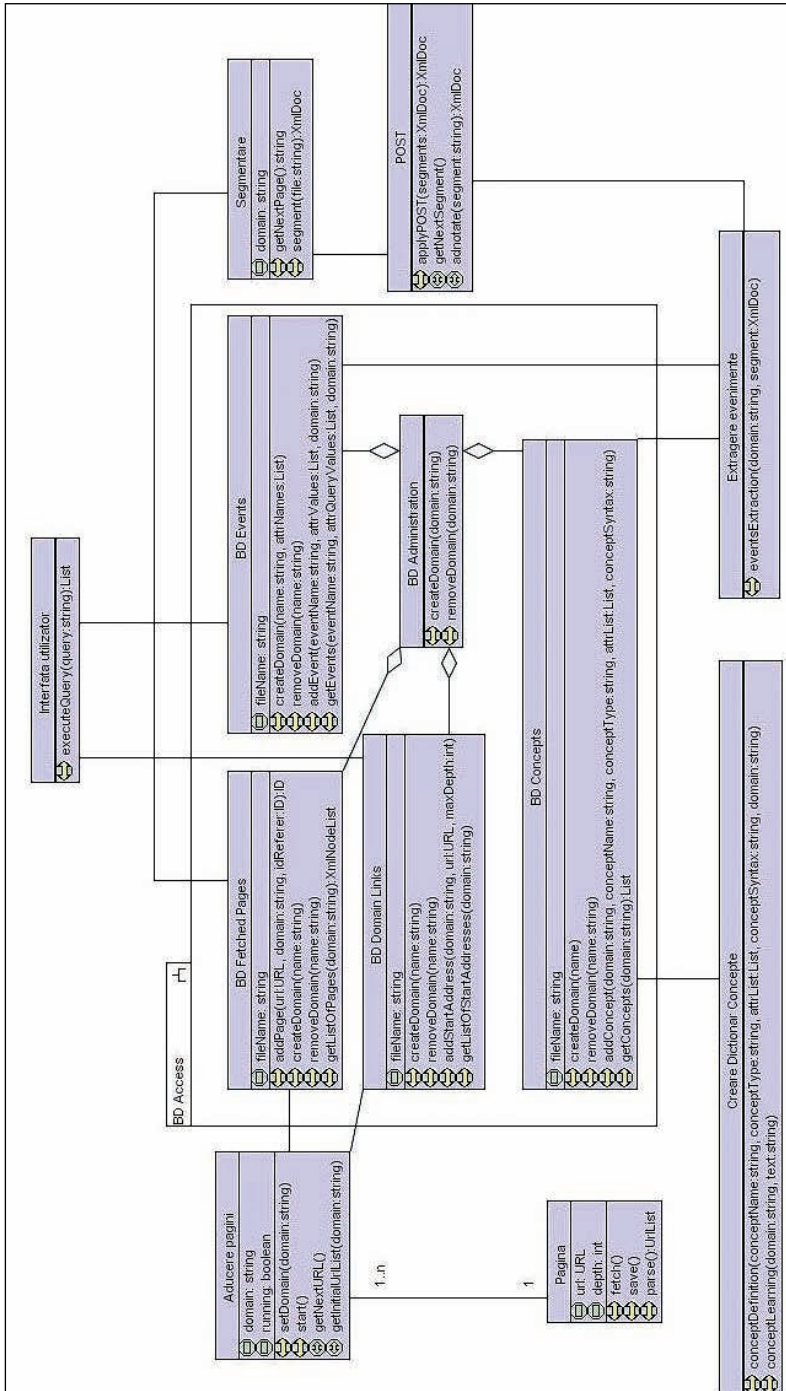


Figure 18: Moduli interface description – classes diagram

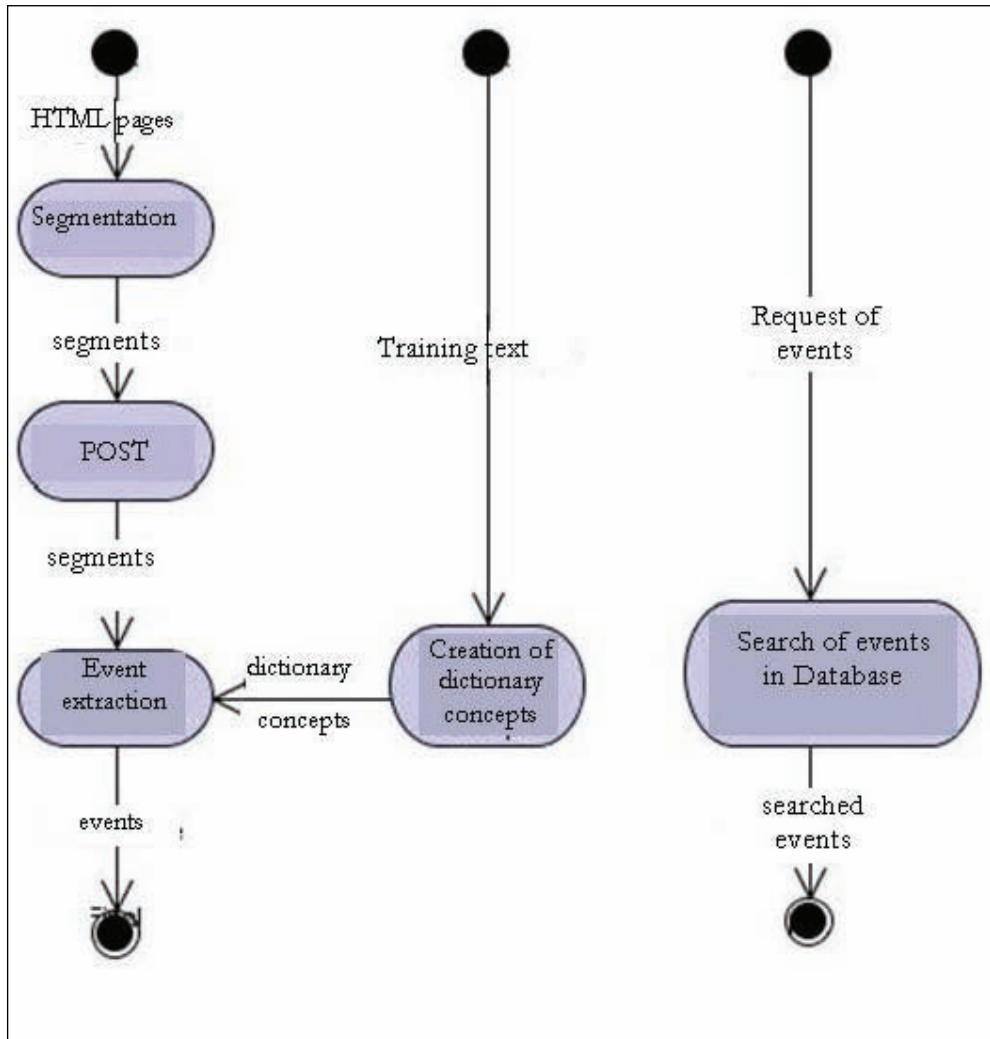


Figure 19: Detailed Description Design

## Detailed description of modules

### Pages Addresses Data Base:

- it corresponds to “BD Domain Links” class
- it is a XML Data Base
- it keeps the list of start pages addresses
- only the start pages are kept in this database, but the fetching module searches and fetches the pages making reference to start pages found on the same server, for a certain fetching depth

```
<xml>
  <domain name=”_tiin_a inform_rii”>
    <urlname=”www.yahoo.com/information science” adancime=3>
    <url name=”www.cnn.com/ information science ” adancime=2>
      ....
  </domain>
</xml>
```

### Pages Data Base:

- it corresponds to BD Fetched Pages class
- it is a XML Data Base
- it keeps the index of fetched pages in XML format, and the pages in distinct files
- only the pages containing text are fetched and kept, not those indicated by links to images or to other insignificant types for application.
- For each page there is an index in the Pages Data Base that contains information about the page url, the last updating date, the name of the file where it was stored, a unique id of each page and the list of pages ids with which a page is in relation.

Ex :

```
<xml>
  <domain name=”_tiin_a inform_rii”>
    <pagina url=”www.yahoo.com/information science /” lastUpdated=”22.11.2002”
      fileName=”information science_com_yahoo_www” id=1>
  </pagina>
  <pagina url=”www.yahoo.com/_tiin_a inform_rii/europe”
    lastUpdated=”22.11.2002”
    fileName=”europe__tiin_a inform_rii_com_yahoo_www” id=2>
  <upPage>1</upPage>
  </pagina>
</domain>
<xml>
```

## **Events Data base:**

- it corresponds to BD Events class
- it is a XML Data Base
- it keeps the identified events, indexed according fields, time locations and other criteria
- Concept = assembly of all items of information (events) of a certain type that follows a certain syntax imposed by concept (ex: {Bucuresti, curriculum}, {information science, Brasov, University, {information science, X, Y}}, for any X and Y valid, plus a series of syntaxes valid for the concept of information science and attributes X and Y)
- The following items of information must be kept for the concept:
- Name of concept
- Type of concept
- List of attributes
- Syntax of concept – positioning the attributes related to a certain concept

## **Web pages fetching:**

- It corresponds to “Pages fetching” class
- It fetches Web pages from Pages Addresses Data Base and it stores them in the Pages Data Base. This module may recursively fetch the pages indicated by links in the current page, from the same server, till a certain fetching depth.
- This module appeals to an external application, specialized on web pages recursive fetching; yet, it can fulfill this function internally.
- It contains analysis function of the page to extract the new links, plus functions of information removal from page (ex: scripts) that are not relevant for the application

## **Segmentation:**

- It corresponds to Segmentation class
- It divides the text from the Web pages within the Pages Data Base in segments
- A segment is a part of a text that represents a sentence or phrase and can be regarded as an entity of atomic information
- The POST applications accept this type of segments as input data
- The segmentation is made on the basis of HTML tags that helps to their delimitation (ex: <P> <BR> etc.) as well as text delimitation elements.

## **POST:**

- it corresponds to “POST” class.
- this module receives the segments and annotates them with the speech parts corresponding to words in segments.
- The module interfaces the program with a Part of Speech Tagger application, which receives a segment and annotates it.

## **Events extraction**

- it corresponds to “Events extraction” class.
- it extracts events from segments annotated, received from the POST module, on the basis of concepts in the Concept-based Dictionary and of WordNet, adding them to the Events Data Base.
- This module is one of the most complex in the system, together with the Concept-based Dictionary and Segmentation Modules.
- A concept is made up of a trigger and a series of attributes that are found as related to the trigger in a certain pattern (that implies a certain syntax).
- It uses algorithms of patterns matching to identify the possible attributes that are subsequently checked by using the WordNet-based information.

## **Concept-based Dictionary**

- it corresponds to Concept-based Dictionary class.
- it creates the concepts specific to a certain field, through learning algorithms, then it adds them to the Concept-based Data Base.
- The Concept-based Dictionary is conceived through either manual identification of representative concepts for a certain field, these being provided to the module to be introduced in the data base, or use of algorithms of learning the concepts on certain learning pages.
- The model with learning algorithms leads to the extraction of concepts that best fit and undertake a good identification of events of similar pages to those from which learning was made.
- The Concept-based Dictionary is one of the most important components within the context of events extraction.

## **User interface:**

- it corresponds to “User interface” class.
- it receives requests for events searching, requests received from the Events Data Base.
- the events are only received from the database, their searching and identification being concluded separately from the Event Extraction module.

## **VI Scientific Research and University Library -proposition of research laboratory for information science-**

University library possesses a significant quantity of data. How shall we benefit from these informational resources? Which is the role of the library?

University library constitutes an integrant part of the didactic and research process. The importance of the library within the didactic process has been exposed and dealt with in many works of specialty. In this day and age the e-learning process and e-book technology have been analyzed and optimized, the library playing a fundamental part.

Which is the role of the library in the scientific research? How does the library contribute to the visibility of the university research? How do we develop the management of the acquisition in order to support the scientific research?

Which are the projects of scientific research we can develop in the field of information science?

Within “Transilvania” University of Brasov we intend to organize a Laboratory of Informational Resources Exploitation within knowledge society.

### **Description of the Proposed Laboratory**

The priority field that the laboratory set of themes affiliates to is that of informative and communication technology.

In the framework of the general direction, the Informational Society of the European Commission, one of the priorities aimed at consists in the increase of the functioning efficacy of the great cultural effects depositaries by means of modern techniques of management and interfacing.

The main objective of the laboratory is the creation of an evaluation and certification structure for the systems of exploitation of the informational resources within the digital library. The object is for this structure to elaborate the validation and verification of some cybermetric, bibliometric and scientometric mechanisms in order to exploit the information in the testing environment: the digital library. The structure created in guise of advanced testing laboratory will generate accredited reference materials for: advanced tests for the exploitation of the information – bibliomining – applied to the management of the digital library collection development; the systemic administration of information; holistic measurements for the library services; visibility of the impact of science in the academic community and economic environment.

The structure that will have been created will be revolutionary, datamining or data exploitation having been successfully used in medical computing, in banking, by the government organizations with a view to combating with crimes,

by ecologists and biologists with a view to discovering the pollution sources and so on and so forth.

For the first time it is created an informational structure which intends to elaborate a set of scientometric, bibliometric and cybermetric indicators, scientifically elaborated and tested, which will be used for research in information science.

The interweaving of the 5 fields – biblioteconomy, sociology, linguistics, engineering and computing – and the cooperation among the researchers will stimulate the accessibility, the information circuit for the knowledge economy and the knowledge society will provide new research possibilities if the access is based upon content and semantics, upon extraction and indexation of the information.

**LERISC intends** to elaborate at a national level informational models, indicators, indicators correlated to the following goals: Development of the bibliomining model. Theoretical concept. Applicative model; Information warehouse (CSI); Statistic models applied to bibliomining; Training of qualified specialists in evaluating, auditing and certifying the conformity for the informational products.

**The objective consists in** the researchers' regrouping, in the reunion of the competencies and in the enhancement of the resources with a view to stimulating the research in the information and communication sciences, with a view to extending their visibility and development and to the elaboration of cybermetric, bibliometric and scientometric indicators.

Within the informational society, the concepts of information and communication have acquired a new dimension: information and communication constitute an epistemological and pragmatic continuum which different specific research is placed upon.

**LERISC intends** to work upon the instruments and methods of access to the information, upon the consequences of the information and communication technologies with respect to the way society goes and to the new forms of sociability and to the generation of an accredited structure in the national information system which should generate informational models.

In **Romania** the reference points and the standards which will constitute the goal of LERISC are currently accomplished in a random and artisan manner, bibliometric and scientometric studies are resorted to without a previous analysis, verification and certification. It is necessary for these operations to be accomplished in an organized framework, in compliance with the European standards, a fact which will ensure the integration of the structure conceived to this purpose in the thematic area according to PC7 and to the goals of the access to the information, dissemination of the information, improvement of the



scientific research and visibility of the Romanian research.

The laboratory will stand for the first research ring chain in this field, research methodologies and approach and analysis proceedings will be elaborated with a view to creating the data warehouse for bibliomining. The development of the knowledge representation models extracted from the textual data, the automatic processing of the textual information remains a major challenge for informational society. The indicators elaborated through these techniques will be used either with a view to administrating the scientific research promotion (bibliometrics and scientometrics), or with technological and concurrential supervision (bibliometrics traditionally applied or bibliomining applied to the semantic web).

### **Scientific Presentation of the Laboratory**

From the presentation of the existing situation it may be noticed that the goal proposed covers an area of preoccupations which IS NOT CURRENTLY COVERED by other similar preoccupations in the country.

University libraries possess integrated library systems. The Library of Transilvania University possesses the ALICE soft, which generate the data base of the library and the on-line catalogue. There has been proposed the acquisition of the LIBERTY soft, which constitutes the superior variant of the ALICE programme, which includes a full-text browser, which allows a total personalization, both introducing the data and for typing the reports and statistics, which possesses complex security rules, which may be administered and used through the intermediary of a browser, from any computer with access to the Internet, anywhere in the world, on the basis of an individual account.

By means of datamining there will be elaborated advanced tests applied to the resources: the database of the publications, the interlibrary loan, the number of the publications that have been accessed, the frequency of apparition of the works that have been accessed, the database of the users etc.

Datamining stands for the automatic practice of search for patterns in big data warehouses, for the extraction of non-trivial information, not previously known and potentially useful from the data, the science of useful information extraction from amounts of data.

The directions of research in exploiting these data are:

- Discovery of associations among objects;
- Grouping of the objects in sets of similar objects (clustering);
- Classification of the objects on the basis of their properties;
- Evaluation of the interest of the facts and properties that have been discovered;
- Preparation of the data (purification, discretization etc)

The practical competencies necessary to the team of researchers who are involved in the accomplishment of the project:

Relational data bases, SQL and its use in C++, JAVA and other languages; Algorithms which work with a variety of data structures; Administration of the data warehouses.

Theoretical models in the field of mathematics which will be used: Clustering: metric spaces, linear algebra and functional analysis; Classification: theory of information, graphs; Lattice rules, association; Theory of information; Probabilities, statistics. The notion of clustering is important for the condensation of data (concise presentation of data), identification of the tendencies within data.

In order to settle the indicators deemed to be obtained by LERISC several algorithms will be made use of:

**1.** Incremental clustering which is characterized by the formation of groups, successively adding objects to the groups (clusters) or forming new groups; **2.** Clustering as partitions, aiming at the determination of the groupings from their connections with the partitions induced by the attributions; **3.** AMICA (A Metric Incremental Clustering Algorithm); **4.** Decision trees.

Techniques resorted to with a view to obtaining the indicators which make the object of LERISC laboratory: Rules of induction; Neuronal networks; Conceptual clustering; Associative discovery;

After the optimization there will be elaborated the model of bibliomining and the indicators of performance proposed for the testing.

- There will be applied measures of cumulative assessment of the library information services. The measurements will be characterized by a holistic perspective, the variables resorted to being: the perspective of the measurement and the subject of the measurement.
- There will be approached the data bases of the library collection and of the users in the manner of the bibliomining archaeological model. There will be generated matrices and indices for the identification of the opportunities and for the orientation of the scientific research.
- There will also be approached the web space through the advanced retrieval system of the information, through statistic and lexical approaches of the textual data.
- There will be elaborated a model of representation of the information extracted from the textual data, which will comprise elements of semantics for the management of the scientific research prognosis (bibliomining, bibliometrics, scientometrics) and a concurrential technological supervision (bibliomining applied to the patent of information).

The generated selective pattern will be characterized by retrieval indicators advanced for testing.

In this respect there will be acquired: **the soft e-Reference** represents an interrogation archive for reference demands placed by the users of the system. The users may search full text in the “knowledge archive” created this way and they may be granted access, according to the security level of the demand, to the answer-solution offered by the reference librarian. This application may be used together with Liberty3 (file of common users) or independently; **the soft Z39.50** which allows “web cataloguing” resorting to any compatible Z39.50 catalogue in the world; the selected records are transferred in Liberty3 by a single click; **LIRIX** – system of mono and multi-linguistic information retrieval; **SPSS** may rapidly generate information of use in the decisional process and offers the possibility of presenting the results with high quality charts and of communicating the results by a variety of reports, inclusively by publication on website. All these empower us to take intelligent decisions, to rapidly discover key factors, patterns and trends in the data. SPSS is used for datamining and for the analysis of the data bases, for marketing studies and any kind of research, as SPSS is the best software in solving business and research problems resorting to statistics. SPSS is a modular line for products completely integrated for the analytical process – planning, data collection, access, data preparation and management, analysis, report drawing up and result presentation. The graphic interface makes it easy to use and it offers all the methods of data management, analysis and presentation in reports you need in order to accomplish even the most intricate analyses.

**Clementine 10.**

The data used in bibliomining will be collected through the multimedia stations processed by the project team in a network of 15 computers.

For the dissemination and evaluation of the proposed indicators there will be acquired a digital type machine XEROX 4110. The typed products on digital copying support and media effects have been accomplished with the copying/printing 4001 machine from XEROX products. They offer an exceptional productivity, reliability and superior flexibility. Media effects being placed at the user's disposal, he is offered the flexibility to create a wide range of printing applications and finite products, such as: brochures, reports, textbooks etc.

After the experimental verification, all indicators will be proposed for accreditation and validation.

The control or the efficacy will be approved after the verification of the impact in the academic environment, library management, economic environment through the sociologic studies and methods.

## Motivation of the Proposed Laboratory

University Library constitutes a part of the national system of libraries, which according to the Law of the libraries no. 334/2002, art. 9, paragraph (1), makes an integrant part of the national informational system according to this article of law, the national library system aims at the *coordinated accomplishment of the national and international library loan, systemic management of the information, accomplishment of the national Catalogue under partition and of the virtual national Library*. In this context, the library, standing for an information centre, has to conceive systems and methods for automatic indexation, search and retrieval of the information. The library has to be oriented towards recordings of digital information which allow mutual information exchanges and their efficient enhancement.

In the framework of the information and documentation services, there has emerged the necessity to dispose of performant indicators that should assess the quality and efficacy of the information comprised in data bases of the activity carried out. The standard SR ISO 11620:2002 – Information and documentation. Performance indicators for libraries – settles a number of assessment indicators for the evaluation of the library activity results, stating for every indicator the objective, the field of application, the definition of the indicator, the methodology, the interpretation of the data and the factors which influence the indicator, the correlated sources and indicators.

The project is based upon an attentive study of the standard SR ISO 11620:2002 which, in the last but one paragraph of the Introduction chapter, page 3, specifies: *There are certain library activities and services for which – during the period this standard was being elaborated – there was noticed the absence of accredited and well documented indicators. This fact is valid for the information services, for the users' training and for the automated services in general. (...). The library and informational community is required to conceive the mechanisms necessary for the elaboration of adequate indicators for these aspects, paying a priority attention to this problem.*

The project intends to realize the laboratory LERISC laboratory with responsibilities at a national level in elaborating **informational models** used in the information and documentation system, which aims at the following essential aspects of the information activity: analysis and representation of the information; organization and depositing within the memory of the information; putting in standard format of the information; on-line search of the information; optimization of the search for information; testing of the interrogation interfaces and of the interrogation instruments; proposition of the assistance instruments of the search for information; accomplishment of the complex and evolutive search strategies; definition of an answer policy to the requirements of search

for information; elaboration of the methods of evaluation of the systems of search for information; identification and validation of the information sources; improvement of the functioning and administration of the information and documentation services; justification of the budgetary allowances at the disposal of the information and documentation services; evaluation of the services with direct access to the information; integration of the elements of added value; evaluation of the qualitative indicators of the provided service.

The definition of the scientometric indicators with respect to the impact of scientific information upon the users' information requirements results from the activities carried out within the information and documentation services, out of whom we can enumerate: the number of the information demands and of the accessed publications; distribution upon fields of scientific interest of the publications and of the information; the necessary of information for different disciplinary and interdisciplinary fields; orientation upon sets of themes in different fields of specialization; quality of data bases and of information sources; influence of the economic and social factors upon the systems of scientific research; the time required for processing the information demand.

In the context of the management of the information from the data bases, there can be established the deficiencies and the necessary of information for different disciplinary and interdisciplinary fields, there can be assured the quality of the development of the data bases and there can be created data bases upon scientific specialization.

LERISC Laboratory proposed for coming into being:

- is vouched for and has the possibility to educate in the spirit of the European norms and to create specialists in the field of certification of products and services;
- will pursue with coherence and consequence the process of harmonization with the European directives and norms;
- has in view a valid and credible system upon which the evaluation of the products credibility should be based;

In the activity of the information centers, in the field approached by the project, the efficacy aims at actions such as systemic management of the information. It is consequently necessary to establish the system of indicators, accredited and well documented, for the evaluation of the information management which should meet the criteria and uses SR ISO 11620 and the European community requirements.

**The risk** in the activity of LERISC laboratory refers to the probability of non-observance of the pre-established goals in terms of **performance** (non-accomplishment of the quality standards), **program** (non-compliance with the execution delay), and cost (exceeding of the budget).

**The management of the risk** stands for a cyclic process, with several distinct phases: identification of the risk, analysis of the risk and reaction to the risk. The managerial team takes into consideration both internal and external risks. The internal risks represent risks that the managerial team may control or influence, while the external risks are beyond its control.

This way there are drawn up control lists which comprise potential sources of internal risk: expected results, staff, modifications of the goals, errors and omissions in design and execution, assessments of the costs and of the execution delays etc.; Moreover, the staff of the laboratory is invited to a formal meeting of identification of the risks. An efficient communication is one of the best sources of risk diminution and execution.

As for the external risk, there are identified the risks imposed from the exterior (through legislation, changes in economics, technology, relation with the trade-unions) by appointing a person who should take part in the meetings of the professional associations, in conferences, and who should run through the publications in the field.

The hypotheses and the risks which may affect the viability of the project are:

- accentuated modification of the currency;
- delays in the payment of the services and in the acquisition of the equipment;
- young employees with small wages may leave the organization;
- the probability of non-compliance with the pre-established goals in terms of performance.

LERISC – the laboratory we propose, which will become a structure of advanced tests for the manipulation of the sets of data for relevant information in the fields of the information science, will comply with the existing standards at an international level and those adopted at a national level.

**The beneficiaries** of LERISC activity will be: University libraries, component parts of the didactic process and of the national system of scientific research; Universities, by the improvement of the training for research and by the visibility of research and access to relevant information; Centers of information-documentation and research; All the users of the library; Researchers and professorial staff in all the fields and especially those in the information science; Managers of the info-documentary systems; Doctoral students and Masteral students in the Universities, for the finalization of their doctoral theses and of their dissertations; the Economic environment through the dissemination of the results by sociologic methods.

Within LERISC there will be determined the necessary mechanisms for the elaboration of accredited indicators, of use in the improvement of library

collection management, in the fluidization and filtration of the informational content and in the retrieval of the relevant information through a model based upon systems of interrogation of the users. By LERISC there will be paved the way for the knowledge society.

### **Technical, economic and social impact**

Within modern society information mean power. And power is knowledge. Within contemporary society not only science fields, but also human being are objects of knowledge. This project-program aims at producing effect both upon human resources and upon fields of research, economic or technical. The impact upon the human resources materializes in the creation of new posts, in the creation of new specialists, of a new job.

The impact upon education and research materializes by the opportunities for education and advanced learning. The most recent discoveries, the latest research will be placed at the disposal of the researchers, of the students, of the masteral students. Within this laboratory of data mining and information mining there will be processed the required information, there will be exploited data bases in the shortest delay.

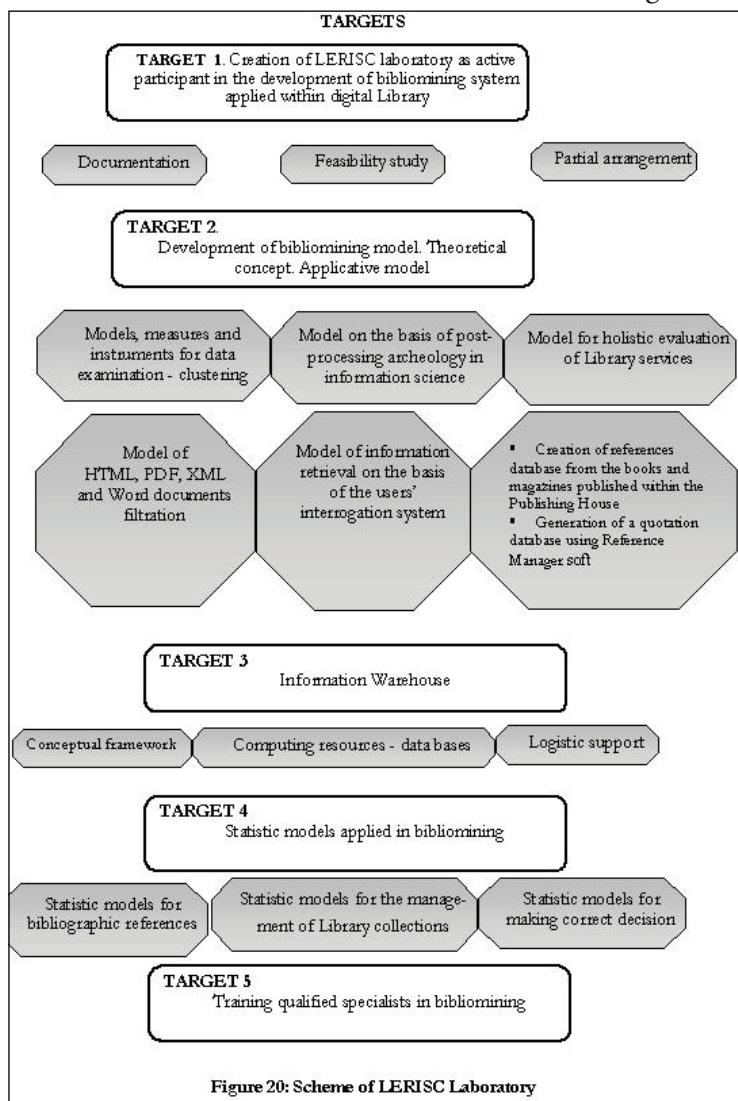
Moreover, by means of the technology and of the created system, we place ourselves at the disposal of the economic-scientific-technical environment. In a world of speed, of information avalanche, of the need of continual evaluation of the development market in any field, we must find solutions in no time. This way, it will be possible to carry out complex evaluations, statistics, to extract and index, by computing, data from the fields of interest for a time economy and for a smaller effort of administration of the great quantity of retrieved documents. This is the reason why the group of professionals who initiate this project, librarians, specialists in information and documentation, sociologists, linguists and engineers, lay the bases of a system which, besides being purely technical, is also human, and addresses people, those who need information. They will train the best infomanagers and archeologists of the web space and will form together a complex team of informational development. Specialists in information and documentation contribute essentially to the selection of the necessary information; Sociologists identify the typologies of persons and establish the categories of users according to their needs and identify the information demands of the academic and economic environment; Linguists interpret the content of the document, extract the syntactic and semantic information from the text of the document and use the information in order to harmonize them with the information necessary to the user; Engineers create the technology so that all specialists' activity should interpenetrate on the waves of the new

communication technologies; Computer scientists design and improve the logistic support, the soft for the exploitation and evaluation of the information.

## TARGETS

### Glossary of terms related to Bibliomining

These definitions are presented only in the context of this tutorial. Many of these terms have other definitions or more refined meanings not represented here.





**AGGREGATION** - An aggregation is used here to mean some way of combining underlying measures to create a single measure. This could be through adding a set of values (number of visits in the last year), dividing one measure by another (average items examined per search), or some other way of combining a number of underlying measures. Typically, this is done for large numbers of library users in order to provide a single measure for evaluation. The disadvantage of creating these aggregations over all library users is that it masks underlying patterns between user communities.

**BIAS** - A bias occurs when decisions are made regarding data collection, aggregation, data cleaning, or other steps in the bibliomining process that create a sample that is not representative of the overall population. One example is a selection bias, where individuals targeted for a user survey are not randomly selected from the population, but rather are selected because they were visiting a library or were part of some subgroup accessible to the researcher. Another example is discarding records that have a missing value for a specific measure in order to calculate an aggregation. Few studies have no bias, but one goal of research is to reduce or control for bias.

**BIBLIOMETRICS** - Bibliometrics is the study of patterns in formal scholarly communication. Common items examined for bibliometrics include citations between works, patterns of authorship, journals of publication, and words used in documents. It is related to informetrics, which is the examination of patterns in a larger body of communication, and webometrics, which explores patterns in Web communication.

**BIBLIOMINING** - Bibliomining is the combination of data warehousing, data mining, bibliometrics, statistics, and reporting tools used to extract patterns of behavior-based artifacts from library systems. The name comes from the combination of the terms “bibliometrics” and “data mining” as bibliomining combines patterns from the creation process and patterns in the use of library services.

**CORRELATION** - Correlation is an indication of the relationship between two measures, and is usually measured between -1 and 1. A high positive correlation (closer to 1) means that when one measure goes up, another one goes up. A negative correlation (closer to -1) means that when one goes up, the other goes down. A correlation near 0 means that there's little observable relationship between the two measures. For example, the number of times someone visits the library should be positively correlated with the number of items that the person checks out. Correlation is useful in prediction and is the basis for many statistical procedures.

**DATA MINING** - Data mining is the exploration of a large cleaned data set using statistical and artificial intelligence methods. Data mining can be

directed (with a specific topic area or goal) or undirected (with no specific goal). The purpose of data mining is to discover patterns that are novel, meaningful, and actionable. In order to determine appropriate patterns, a domain expert needs to be involved with the data mining process.

**DATA WAREHOUSE** - A data warehouse is a secondary copy of operational data reformatted for analysis and reporting. Typically, the data warehouse will connect data from different parts of an operational system. In addition, the data warehouse can be a place for external data to be connected to operational data. Creating a data warehouse allows the time spent gathering and cleaning data for one project to be easily used in other projects.

**DEMOGRAPHIC SURROGATE** - A demographic surrogate is a set of variables used to replace the personally identifiable information about a patron. The demographic variables selected should be useful to the library in decision-making and justification. The result is that these surrogates can help the librarian to better understand different communities who use library services.

**DOMAIN EXPERT** - In the setting of a particular library, the domain expert is someone who has worked in that library for a significant amount of time. The patterns found with data mining should correspond to the observations of the domain expert.

**EVIDENCE-BASED LIBRARIANSHIP** - Evidence-Based Librarianship (EBL) was originally developed from concepts used in Evidence-Based Medicine. In traditional EBL, the librarian starts by gathering research pieces that address a particular topic. The results from these pieces are then combined to create Evidence, which is then used to resolve the decision-making need. Bibliomining concepts can be used to create a different type of Evidence by collecting data from many different libraries in the same standard format. This data warehouse can then be tapped as needed for Evidence.

**ONLINE ANALYTICAL PROCESSING (OLAP)** - An Online Analytical Processing tool allows the exploration of a dataset through interactive manipulation of different variables over selected time periods. The resulting report can be refined through the expansion or collapse of variable categories. In regard to bibliomining, an OLAP tool gives a staff member with little training the ability to explore the library's data warehouse as needed.

**OUTCOME-BASED EVALUATION** - The goal of outcome-based evaluation (OBE) is to demonstrate the impact of library services. OBE starts by the selection of measurable outcomes that the library should bring about. Measures are then selected and gathered based upon these outcomes. The measures are then evaluated to determine if an impact has been made. The result is a guided evaluation plan that allows the librarians to understand what types of impact their library services are having.

**OUTLIER** - An outlier is a data point that is either significantly higher or lower than the bulk of the population. Many statistical methods are affected by outliers. For example, an average salary can be skewed by a small number of outliers with very high salaries. The resulting average would not provide a reasonable understanding of the desired measure.

**STANDARD DEVIATION** - The standard deviation of a measure indicates how spread out the individual data points are. A small standard deviation (when compared to the average) indicates that most of the individual data points are close to the average. A large standard deviation means that the average may not be very useful in understanding the data set.

**STOVEPIPING** - Many operational library systems keep the data for different parts of the operation separate. For example, the Interlibrary Loan data, Circulation data, and E-Resource use data usually live in three different systems. This concept, called stovepiping, makes it difficult to create library studies that look at the use of material across different systems. A data warehouse is a place where data from these different systems can be matched and formatted for analysis.

**VISUALIZATION** - Visualization techniques take a complex data set and attempt to make it more understandable by using graphics to display components of the data set. Graphs and charts allow for the identification of trends and comparisons of many different data points. Graphical webs showing connections between works can be used to demonstrate links or citations between works or authors. Since visualization simplifies a complex data set, it also can mask information or mislead the viewer in their interpretation of the data.



## REFERENCES

1. Agent Technology: Foundations, Applications and Markets. Edited by Jennings, N.R.; Wooldridge, M.J. Berlin: Springer Verlag, 1998.
2. American Civil Liberties Union. (n.d.). USA PATRIOT act. Retrieved May 18, 2004 from <http://www.aclu.org/SafeandFree/SafeandFree.cfm?ID=12126&c=207>
3. BĂLĂIȚĂ, Rareș, HULEA, Mircea, OLARIU Cristian. Data Mining. Inteligență artificială – Referat, Iași: Universitatea Tehnică “Gh. Asachi” Iași, Facultatea de automatică și calculatoare, 2002/2003
4. BALL, Philip. One-page book on horizon, Nature - Science update (on line), Tuesday 24 April 2001 ('Paper could soon be obsolete. Researchers from Bell Laboratories in New Jersey and the E Ink Corporation in Cambridge, Massachusetts, have found an economical way to make sheets of flexible plastic that can be electronically activated to display different pages of black and white lettering. A single sheet of this 'electronic paper' could, in principle, contain an entire book. Information, held in some portable memory device such as a CD player, could be fed onto the plastic page to appear as black and white characters or images. Each page of information could be called up in sequence, whereupon the 'electronic ink' on the sheet rearranges itself into a new configuration.')
5. BANERJEE, K. Is data mining right for your library? Computers in Libraries, 1998, 18(10), 28-31.
6. BARABÁSI, A. Linked. New York: Penguin Group, 2003
7. BERRY, J.; LINOFF, G. Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management. 2nd edition. Indianapolis, IN: Wiley, 2004
8. BOLLEN, J.; LUCE, R.; VEMULAPALLI, S.; XU, W. Usage analysis for the identification of research trends in digital libraries. D-Lib Magazine, 2003, 9(5). Available online at <http://www.dlib.org/dlib/may03/bollen/05bollen.html>.
9. BORGMAN, C. L.; FURNER, J. Scholarly communication and bibliometrics. In B. Cronin (Ed.), Annual Review of Information Science and Technology (Vol. 36, pp. 3-72). Medford, NJ: Information Today, 2002.
10. BORGMAN, C. L. (Ed.) Scholarly Communication and Bibliometrics, Newbury Park, CA: Sage Publications, Inc., 1990.
11. BÖRNER, K.; CHEN, C.M.; BOYACK, K.W. Visualizing knowledge domains. Annual Review of Information Science and Technology 37, 2003, 179-255.
12. BOULTON, Richard E.S; LIBERT, Barry D.; SAMEK, Steve M. Cracking the Value Code. How Successful Businesses Are Creating Wealth. In the New Economy, Harper Business, New York, 2000.
13. BRADSHAW, J. Ed., Software Agents, Menlo Park, CA, AAAI Press, 1997.

14. BROWN, John Seely; DUGUID, Paul. Organizing knowledge. In: California Management Review, 40(3): 90-111, 1998 Spring.
15. BUCKLAND, M. Five grand challenges for library research. In: Library Trends, 2003, 51(4), p. 675-686.
16. CANTU-PAZ, E.; CHANDRIKA, Kamath. On the Use of Evolutionary Algorithms. In: Data Mining; <http://www.llnl.gov/CASC/sapphire/pubs/heuristic.pdf>
17. CHAUDHURI, S.; DAYAL, U. An overview of data warehousing and OLAP technology. In: SIGMOD Record, 1997, 26(1), 65-74.
18. CHEESEMAN, P.; J. STUTZ. Bayesian Classification: Theory and results. Eds. Fayyad, U., Piatetsky-Shapiro, G, Smyth, P. and Uthurusamy, R. In: Advances in Knowledge Discovery and Data Mining, Cambridge, MA: AAAI/MIT Press, 1996.
19. CHEN, C. Visualising semantic spaces and author co-citation networks in digital libraries. In: Information Processing and Management 35(3), 1999, p.401-420.
20. CHIP, Elliott. Building the Wireless Internet. In: IEEE Spectrum, January 2001, p. 14-16.
21. COLLIER, K. B.; CAREY, E.; GRUSY, C.; MARJANIEMI, D; SAUTTER, D. A Perspective on data Mining, Northern Arizona University, July 1998, [insight.nau.edu/downloads/DM%20Perspective%20v2.pdf](http://insight.nau.edu/downloads/DM%20Perspective%20v2.pdf)
22. Commission of the European Communities, Europe 2002 – Impact and Priorities, A communication to the Spring European Council in Stockholm, 23- 24 March 2001, COM (2001) 140 final, Brussels, 13.3.2001.
23. Competing visions. In “Progress Toward Digital Libraries”, eds. Christine L. Borgman, Gary Marchionini and Edward A. Fox, Special Issue, Information Processing & Management, 35(3), May, 1999, p. 227-244,
24. DAN Nica, Guvern, Cetățean, Societate informațională, Editura SEMNE, 2001.
25. DAVENPORT, Thomas D.; PRUSAK, Laurence. Working knowledge: How Organizations Manage What They Know, Harvard Business School Press, 1998.
26. DRĂGĂNESCU, M. Procesarea mentală a informației, cap.5: Psihologia agenților inteligenți. In Memoriile Secțiilor științifice, Academia Română, Seria IV, Tomul XX, 1997, p. 270-272,
27. DRĂGĂNESCU, M. De la societatea Informațională la societatea cunoașterii, Editura Tehnică, București, 1991, p.237 [www.ici.ro/cărți.html](http://www.ici.ro/cărți.html)
28. DRĂGĂNESCU, Mihai. Conștiința, frontieră a științei, frontieră a omenirii, comunicare la sesiunea Comitetului Român pentru Istoria și Filosofia Științei și Tehnicii, Academia Română, 18 octombrie 2000. In: Revista de filosofie, XLVII, Nr.1-2, 2000, p. 15-22.
29. DRĂGĂNESCU, Mihai. Discurs la Consfătuirea Președintelui României cu manageri și specialiști din domeniul Tehnologiei Informației, București,

- Palatul Cotroceni, 12 aprilie 2001.
30. DRĂGĂNESCU, Mihai. Societatea conștiinței – cuvânt la Fundația Henri Coandă, Atheneul Român, 30 iunie 2001.
  31. DRĂGĂNESCU, Mihai. Societatea cunoașterii și cartea electronică, comunicare. Simpozionul Cartea Electronică, Academia Română, 21 iunie 2001.
  32. DRĂGĂNESCU, Mihai. The Economy of the Knowledge Society – intervenție la Conferința Noua economie - o șansă pentru România, Bucuresti, Hotel Marriott, 20 martie 2001.
  33. DRĂGĂNESCU, Mihai. Conștiința, frontieră a științei, frontieră a omenirii, comunicare la sesiunea CRIFST, Academia Română. 18 oct. 2000. In: Revista de filosofie, XLVII, Nr. 1-2, 2000, p.15-22.
  34. DRĂGĂNESCU, Mihai. [www.racai.ro/~dragam](http://www.racai.ro/~dragam).
  35. DREXLER, K.Eric. Nanosystems: molecular machinery, manufacturing and computation. New York, NY: John Wiley, 1992.
  36. DRUCKER, Peter F. Beyond the Information Revolution, The Atlantic Monthly, Digital Edition, <http://www.theatlantic.com/issues/99oct/9910drucker3.htm>
  37. Data Mining, Turn computer loose on your data and you don't know what they'll come up with--that's the whole point, coordonator E.X. de Jesus, In: BYTE, October, 1995. <http://www.byte.com/art9510/sec8/sec8.html>
  38. EIRINAKI, M.; VAZIRGIANNIS, M. Web mining for web personalization. ACM Transactions on Internet Technology, 2003, 3(1), p.1-27.
  39. EROIL Gelenbe; ŞEREF Esin; XU Yhiguang, Simulation with Learning Agents, p.148-157.
  40. European Commission Enterprise DG - European Observatory on Intangible Assets, Policy Trends in Intangible Assets, 8 Nov. 2000 (prin Internet, <http://www.eu-intangibles.net>).
  41. European Commission, Community research, ISTAG, Scenarios for Ambient Intelligence in 2010, Final Report, February 2001, Seville.
  42. FAYYAD, U.M.; PIATETSKI-SHAPIRO, G.; SMYTH P.; UTHURUSAMY R. Advances in Knowledge Discovery and Data Mining, AAAI/MIT Press, 1996.
  43. FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. From Data Mining to Knowledge Discovery in Data bases. In: AI Magazine, Fall 1996a.
  44. FAYYAD, U. Editorial. Data Mining and Knowledge Discovery, 1997.
  45. FAYYAD, USAMA; PIATETSKY-SHAPIRO, G.; SMYTH, P. From Data Mining to Knowledge Discovery: An Overview: In: Advances in Knowledge Discovery and Data Mining, AAAI/MIT Press, 1996b.
  46. FAYYAD, USAMA; PIATETSKY-SHAPIRO, G.; SMYTH, P. Knowledge Discovery and Data Mining: Towards a Unifying Framework. In: KDD-96 Conference Proceedings. Ed. E. Simoudis, J. Han, and U. Fayyad, AAAI

- Press, 1996c.
47. FAYYAD, Usana; UTHUTASAMY, Ramasamy. Data Mining and Knowledge Discovering in Data bases. In: Comm. of ACM, November 1996, vol.39, nr.11, p.25-26.
  48. FRAWLEY, W.; PIATETSKY-SHAPIRO, G.; MATHEUS; C. Knowledge Discovery in Data bases: An Overview. In: AI Magazine, Fall 1992, pp 213-228.
  49. GARDNER, Claudia. IBM Data Mining Technology, 11.04.96.  
<http://booksrv2.raleigh.ibm.com/cgi-bin/book.mgo.cmd/BOOKS/datamine> sau <http://www.almaden.ibm.com/cgi-bin/stss/data-mi>
  50. GEPPERT, Linda. The New Chips on the Block. In: IEEE Spectrum, January 2001, p. 66-68.
  51. GEYER-SCHULZ, A.; HAHLER, M.; NEUMANN, A.; THEDE, A. (2003). An integration strategy for distributed recommender services in legacy library systems. Editors: M. Schader, W. Gaul, and M. Vichi, In: Between Data Science and Applied Data Analysis, Proceedings of the 26th Annual Conference of the Gesellschaft für Klassifikation e.V., University of Mannheim, July 22-24, 2002, Studies in Classification, Data Analysis, and Knowledge Organization, pages 412-420. Springer-Verlag, July 2003.
  52. GRAYSON, C. et al. Mining your hidden resources. In: Across the Board, 35(4): 23-28, April 1998.
  53. GREBNER, Ceryl. Excavate Your Data. In: Datamation, May 1, 1996, p.40-41.
  54. GURAN, Marius. Societatea informațională; propuneri pentru o strategie națională – studiu, aprilie 2001, prezentat la ședința comună a Secției pentru știința și Tehnologia Informației și Forumului pentru Societatea Informațională ale Academiei Române.
  55. HAN, J.; SIMOUDIS, E. Proceedings of the First International Conference on Knowledge Discovery and Data Mining (KDD-95), AAAI Press, 1995, <http://www.aaai.org/Publications/Catalog/KDD> și <http://www-aig.jpl.nasa.gov/KDD95>
  56. HAND, D.; MANNILA, H; SMYTH, P. Principles of Data Mining. MIT Press, Cambridge, MA, 2001
  57. HOPKINS, John F, FISHWICK, Paul A. A Three-Dimensional Human Agent Metaphor for Modelling and Simulation, p.131-147.
  58. [http://en.wikipedia.org/wiki/Data\\_Mining](http://en.wikipedia.org/wiki/Data_Mining) (accesat în 10.01.2006 )
  59. <http://info.gte.com/~kdd/index.html> (accesat în 14.02.2006 )
  60. <http://www.angoss.com/> (accesat în 28.01.2006 )
  61. <http://www.anvilinformatics.com/> (accesat în 22.02.2006 )
  62. [http://www.cio.com/archive/051598\\_mining.html](http://www.cio.com/archive/051598_mining.html) (accesat în 10.03.2006 )
  63. <http://www.cognos.com/> (accesat în 23.02.2006 )
  64. <http://www.cs.bham.ac.uk/~ano/TheDataMine.html> (accesat în 19.02.2006 )



65. <http://www.informatik.univ-trier.de/~ley/journals> (accesat în 17.01.2006 )
66. <http://www.microsoft.com/sql/productinfo/datamine.htm> (accesat în 18.02.2006 )
67. <http://www.oracle.com/ip/analyze/warehouse/datamining/> (accesat în 15.01.2006 )
68. [http://www.pilotsw.com/r\\_and\\_t/whatpaper/datamine/](http://www.pilotsw.com/r_and_t/whatpaper/datamine/) (accesat în 24.02.2006 )
69. [http://www.scitegic.com/products\\_services/pipeline\\_pilot.htm](http://www.scitegic.com/products_services/pipeline_pilot.htm) (accesat în 28.02.2006)
70. [http://www.tcs.hut.fi/~helger/crypto/link/data\\_mining](http://www.tcs.hut.fi/~helger/crypto/link/data_mining) (accesat în 05.02.2006 )
71. <http://www.uts.edu.au/fac/hss/Departments/DIS/km/introduct.htm#Char> (accesat în 17.02.2006 )  
<http://www.uts.edu.au/fac/hss/Departments/DIS/km/Resources/KMIntro/index.htm> (accesat în 11.01.2006 )
72. <http://www3.shore.net/~kht/text/dmwhite/dmwhite.htm> (accesat în 12.03.2006 )
73. <http://www-4.ibm.com/software/data/iminer/fordata/about.html> (accesat în 13.02.2006 )
74. <http://www.megaputer.com/dm/index.php3> (accesat în 09.01.2006 )
75. <http://www.std.com/CreativeData/credata/termin.html> (accesat în 15.01.2006 )
76. <http://www.racai.ro/books/doe> (accesat în 30.01.2006 )
77. IANCU, Ștefan. Managementul cunoștințelor, Univers ingineresc, XII, 2001, Nr. 3, 4, 6,7, 8, 10.
78. IEEE Transaction on Knowledge and Data Engeneering, vol.8, nr.6, December 1996, <http://www.computer.org/pubs/fkde/tok.html>, IEEE Expert, October, 1996, <http://www.computer.org/pubs/expert/feature.html>, Communications of ACM, vol.39, 11, November 1996
79. IEEE Transactions on Neural Networks, vol.10, 1999; vol.11, 2000; vol.12, 2001.
80. Information Institute of Syracuse (2004). QABuilder. Retrieved May 21, 2004 from <http://vrd.org/qabuilder.shtml>
81. Interoperability, Scaling, and the Digital Libraries Research Agenda, report of the 1995 IITA DL Workshop
82. IONIȚĂ, Angela. Asupra termenului de minerit de date (Data Mining), Institutul de Cercetare pentru Inteligență Artificială, Academia Română, <http://www.ici.ro/ici/revista>
83. ISO/IEC 13249-6 SQL/MM Data Mining. Edited by Friedemann Schwenkreis, IBM Deutschland Entwicklung GmbH <http://research.microsoft.com/~jamesrh/hpts2001/submissions/FriedemannSchwenkreis.htm>
84. JACKSON, J.E. An User's Guide to Principal Components, New York,

- NY: John Wiley, 1991.
85. JOHNSON, M. Archeological Theory: An Introduction. Oxford: Blackwell, 1999.
  86. KANTE, Rosabeth Moss. Evolve! Succeeding in the Digital Culture of Tomorrow, Harvard Business School Press, Boston, Massachusetts, e-Book MSReader edition, 2001.
  87. KAO, S.; CHANG, H.; LIN, C. Decision support for the academic library acquisition budget allocation via circulation database mining. In: Information Processing & Management, 2003, 39(1), 133-148.
  88. KDD-97, The Third International Conference on Knowledge Discovery and Data Mining, August 14-17, 1997
  89. KOGAN, A.; SUDIT, F.; Vasarhelyi M. The Internet Guide for Accountants: a living book, Prentice Hall, 1997. sau U.Fayyad, G.Piatetsky-Shapiro, P.Smyth, The KDD Process for Extracting Useful Knowledge from Volumes of Data, Comm. of AC M, 39, 11, p.27-34.
  90. KOSTOFF, R., del Río; HUMENIK, J.; GARCÍA, E.; RAMÍREZ, A. Citation Mining: Integrating Text Mining and Bibliometrics for Research User Profiling. In: Journal of the American Society for Information Science and Technology, 2001, 52(13), 1148-1156.
  91. KOULOPOULOS, Thomas M.; SPINELLO, Richard A.; WAYNE, Toms. Corporate Instinct: Building a Knowing Enterprise for the 21st Century, Van Nostrand Reinhold, 1997
  92. KRIVADA, Ceryl D.; Unearthing Underground Story Data, Cover Story, May, 1996, <http://www.lmag.com/9605mine.html>
  93. LEONARD-BARTON, Dorothy. Wellsprings of Knowledge: Building and Sustaining the Sources of Innovation, Harvard Business School Press, 1955.
  94. LIMPAA, H. Privacy-preserving Data Mining. Retrieved May 18, 2004
  95. LOGAN, Brian; THEODOROPOULOS, Georgios. The Distributed Simulation of Multiagent Systems, in Special issue on agents in modelling and simulation: exploiting the metaphor. Eds. A.M. Uhrmacher, P.A. Fishwick, B.P. Zeigler, In: Proceedings of the IEEE, february 2001, p. 174-185.
  96. LUCEY, T. Management Information Systems, DP Publications Ltd., Ed.6, 1991.
  97. MADHAVAN, Ravindranath; GROVER, Rajiv. From embedded knowledge to embodied knowledge: New product development as knowledge management. In: Journal of Marketing. 62(4), 1-12, 1998 Oct..
  98. MAFRUZ ZAMAN, ASHRAFI D.; TANIAR, K.; SMITH, A. A Data Mining Architecture for Clustered Environments, Publisher: Springer-Verlag Heidelberg, ISSN: 0302-9743
  99. MARCHIONINI, Gary; FOX, EDWARD A. "Progress toward digital libraries: augmentation through integration", guest editors' introduction to

- “Progress Toward Digital Libraries”. Eds. Gary Marchionini and Edward A. Fox, Special Issue, *Information Processing & Management*, 35(3), May 1999, pp. 219-225.
100. MARCHIONINI, Gary; FOX, EDWARD A. “Progress toward digital libraries: augmentation through integration”. In: *Information Processing & Management*, 35(3):219-225, May 1999
  101. MARSHALL, Lucy. Facilitating knowledge management and knowledge sharing: New opportunities for information professionals, *Online*. 21(5): 92-98. 1997 Sep/Oct..
  102. MCCLURE, C. Increasing the usefulness of research for library managers: Propositions, issues, and strategies. *Library Trends*, 1989, 38(2), 280-294.
  103. MCCLURE, C.; LANKES, R. D.; GROSS, M.; CHOLTICO-DEVLIN, B. (2002). *Statistics, Measures, and Quality Standards for Assessing Digital Library Services: Guidelines and Procedures*. Syracuse, NY: ERIC Clearinghouse on Information & Technology. Retrieved May 21, 2004 from <http://quartz.syr.edu/quality/>
  104. MENZIES, T.; HU, Y. Data Mining For Very Busy People. In: *IEEE Computer*, October 2003, pp 18-25.
  105. MERKLE, Ralph C. Nanotechnology: What Will It Mean? In: *IEEE Spectrum*, January 2001, p.19-21.
  106. META Group Application Development Strategies. *Data Mining For Data Warehouse: Uncovering for Hidden Patterns*, 13.07.1995 and Pilot Software Group, *Dat Mining - Discovering Hidden Values in Your data Warehouse*, Whitepaper
  107. MICHAELS, James W. How New is the New Economy? In: *Forbes*, October 11, 1999, p.47.
  108. MICHAIL, A. Data mining library reuse patterns in user-selected applications. 14th IEEE International Conference on Automated Software Engineering, Washington, DC. In: *IEEE Computer Society*, 1999, 24-33.
  109. *Mobile Agents '98*, Edited by Rothermel, M.J.; F.Hohl, New York: Springer Verlag, 1998.
  110. MOORE, Samuel K. In: *IEEE Spectrum*, January 2001, p.25.
  111. MURPHY, D.E. Some librarians use shredder to show opposition to new F.B.I. powers. In: *New York Times*, 2003, April 7, pp. A12.
  112. National Academy of Sciences (USA), *Building a Workforce for the Information Economy*, Report, National Academy Press, Washington, D.C., 2001, Internet edition on [http://books.nap.edu/htm/IT\\_workforce](http://books.nap.edu/htm/IT_workforce)
  113. NICHOLSON, S.; STANTON, J. Gaining strategic advantage through bibliominig: Data mining for management decisions in corporate, special, digital, and traditional libraries. Eds. H. Nemati and C. Barko, *Organizational data mining: Leveraging enterprise data resources for optimal performance* (pp.247-262). Hershey, PA: Idea Group Publishing, 2003.

114. NICHOLSON, S. The bibliomining process: Data warehousing and data mining for library decision-making. In: Information Technology and Libraries, 2003, 22 (4), 146-151.
115. NICHOLSON, S. Bibliomining bibliography. The Bibliomining Information Center. Retrieved March 1, 2004a, from <http://bibliomining.org>
116. NICHOLSON, S. A conceptual framework for the holistic measurement and cumulative evaluation of library services. Journal of Documentation, 2004b, 60(2), 162-182.
117. NICHOLSON, S. A framework for Internet archeology: Discovering use patterns in digital library and Web-based information resources. First Monday 10(2). Retrieved March 23, 2005 from [http://www.firstmonday.org/issues/issue10\\_2/nicholson/index.html](http://www.firstmonday.org/issues/issue10_2/nicholson/index.html)
118. NICHOLSON, Scott. The Basis for Bibliomining: Frameworks for Bringing Together Usage-Based Data Mining and Bibliometrics through Data Warehousing in Digital Library Services. In Information Processing and Management
119. NICOLAE, Gabriel; DRAGOMIRESCU, Liviu; BARBOS, Florin; HOBEANU, Claudiu; ICRIVERZI, George. Complemente de informatica
120. NISO. (2004). Z39.7 library statistics – E-metrics data dictionary. Retrieved May 18, 2004 from <http://www.niso.org/emetrics/>
121. NITCHI, S.I.; AVRAM-NITCHI, R. Data Maning, o nouă eră în informatică.
122. NITCHI, S.I.; RACOVITAN, D. – Coordonatori. Bazele prelucrării informațiilor și tehnologie informațională. Ed. Intelcredo, 1996.
123. NONAKA, Ikujiro; HIROTAKA; Takeuchi. The Knowledge Creating Company, Oxford University Press, 1995.
124. PAMBUCCIAN, Varujan. Expunere la lansarea Programului Societatea Informațională – Societatea cunoașterii (SI-SC), Academia Română, 10 aprilie 2001.
125. PATEL, L.; PERRY, S.; TAYLOR, D.; BROWN, A.; TIKE, S. The Future of Data Mining <http://www.aston.ac.uk/~golderpa/CS342/grouppages/dsg9/fut.html>
126. PEIRCE, C. S.: Deduction, induction and hypothesis. În: Popular Science Monthly, 13, 1878.
127. Pilot Software. An Introduction to OLAP, [http://www.pilotsw.com/r\\_and\\_t/whtpaper](http://www.pilotsw.com/r_and_t/whtpaper) sau Creative Data, Data Warehouse Terminology
128. PITAC Report, “Biblioteci digitale: Acces universal la cunoașterea umană”, 2001, p. 3
129. PRODI, Romano. Speech at the German Confederation of Trade Unions’ (DGB) Federal Presidium, Brussels, 7 November 2000, apud Cordis

- Focus, European Commission, 20 November 2000, p.1 and 3.
130. Project Counter. Counter – Counting Online Usage of Networked Electronic Resources, Retrieved May 18, 2004 from <http://www.projectcounter.org>
  131. ROONEY, David; Mandeville, Thomas. The Knowing Nation: A Framework for Public Policy in a Post-industrial Knowledge Economy, In: Prometheus, 1998, 16 (4) pp. 453-467.
  132. QUÉAU, Philippe - Directeur de la division de l'information et de l'informatique de l'Unesco, La nécessaire définition d'un bien public mondial - A qui appartiennent les connaissances? Le monde diplomatique, Janvier 2000, pages 6 et 7.  
<http://www.monde-diplomatique.fr/2000/01/QUEAU/13278.html>
  133. Rensselaer Polytechnic Institute, Data Mining and Knowledge Discovery. <http://www.vpi.edu/~vanepa2/>
  134. RIEZEMAN, Michael J. Optical Nets Brace for Even Heavier Traffic. In: IEEE Spectrum, January 2001, p. 44-46.
  135. ROGERS, J. A. et al. Paper-like electronic displays: large-area rubber-stamped plastic sheets of electronics and microencapsulated electrophoretic inks. In: Proceedings of the National Academy of Sciences USA, 2001, 98, 4835-4840.
  136. SANDSTROM, P.E. Scholarly communication as a socioecological system. In: Scientometrics, 2001, 51(3), 573-605.
  137. SARACEVIC, T.; KANTOR, P. Studying the value of library and information services: Part 1, Establishing a theoretical framework. In: Journal of the American Society for Information Science, 1997, 48(6), 527-542.
  138. SARJOUGHIAN, Hessam S.; YIEGLER, Bernard P.; HALL, Steven B. A Layered Modeling and Simulation Architecture for Agent-Based System Development, in Special issue on agents in modelling and simulation: exploiting the metaphor, eds. A.M. Uhrmacher, P.A. Fishwick, B.P. Zeigler. In: Proceedings of the IEEE, february 2001, p.201-213.
  139. SCHILLER, Dan. Digital Capitalism, MIT Press, Hardcover 1999, Paperback 2000, eBook (MSReader) 2000.
  140. SIMOUDI, E.; HAN, J.W.; FAYYAD; U. Proceedings of the Second International Conference Discovery and Data Mining (KDD-96), AAAI Press, <http://www-aig.org/Publications/Catalog/KDD/>, sau <http://www-aig.jpl.nasa.gov/KDD95/>
  141. Source Book on Digital Libraries. editor Edward A. Fox, 1993, pg. 65
  142. SOUTH, S. Method and Theory in Historical Archeology. New York: Academic Press., 1977
  143. SRIVASTAVA, J.; COOLEY, R.; DESHPANDE, M.; TAN, P.T. Web usage mining: Discovery and applications of usage patterns from Web data. In: SIGKDD Explorations, 2000, (1)2, 12-23.

144. STEWART, Thomas A., *Intellectual Capital: The New Wealth of Organizations*, Currency/Doubleday, 1997
145. STYX, Gary. Trends in Nanotechnology; Waiting for Breakthroughs, *Scientific American*, April 1996, p.78-83.
146. TAIPALE, K. A. Data Mining and Domestic Security: Connecting the Dots to Make Sense of Data (<http://www.stlr.org/cite.cgi?volume=5&article=2>), Center for Advanced Studies in Science and Technology Policy (<http://www.advancedstudies.org/>). 5 Colum. Sci. & Tech. L. Rev. 2 (December 2003).
147. TALPEȘ, Florin - Managerul general Softwin. Inițiativa Open e-book pe plan internațional– comunicare, Simpozionul Cartea Electronică, Academia Română, 21 iunie 2001.
148. TECUCI, Gheorghe. *Building Intelligent Agents, An apprenticeship Multistartegy Learning Theory, Methodology and Case Studies*, Academic Press, 1998.
149. TEODORESCU, Horia-Nicolai; LAKHMI, C. Jain; KANDEL, Abraham. *Hardware Implementation of Intelligent Systems*, Physica-Verlag, Springer-Verlag Company, Heidelberg, 2001.
150. The Corporate KDD Bookmark, <http://www.cs.su.oz.au/~thierry/ckdd.html>
151. The Scope of the Digital Library, Draft Prepared by Barry M. Leiner for the DLib Working Group on Digital Library Metrics, 1998
152. THEARLING, Kurt. From Data Mining to Data Base Marketing, White Paper, 95/02, October 1995, Data Inteligent Group, Pilot Software, <http://www.santafe.edu/~kurt/wp9502.html>
153. WAGNER, Gerd. The AOIS Glossary. On Agent-orientation in Information Systems.
154. WHITE, H.D.; MCCAIN, K.W. Bibliometrics. In M.E. Williams (Ed.) *In: Annual Review of Information Science and Technology 24*. Medford, NJ: Information Today, 1989, p.99-168.
155. WHITE, H.D.; MCCAIN, K.W. Visualizing a discipline: An author co-citation analysis of information science 1972-1995. In: *Journal of the American Society of Information Science*, 49(4), 1998, 327-355.
156. WILKINSON, D.; THELWALL, M.; Li, X. Exploiting hyperlinks to study academic Web use. In: *Social Science Computer Review*, 21(3), 2003, 340-351.
157. WILSON, Linda F. and others, A Framework for Linking Distributed Simulations Using Software Agents, în Special issue on agents in modelling and simulation: exploiting the metaphor, eds. A.M. Uhrmacher, P.A. Fishwick, B.P. Zeigler. In: *Proceedings of the IEEE*, february 2001, p. 186-199.
158. WURSTER, Thomas S.; EVANS; Philip. Blown to bits. How the new economics of information transforms strategy, Harvard Business School

- Press, Boston, Massachusetts, Hardcover, 2000. Microsoft Reader edition (eBook), 2000.
159. WURSTER, Thomas S.; EVANS, Philip. Strategy and the New Economics of Information, Harvard Business Review, Sept.-Oct. 1997.
  160. ZUCCA, J. Traces in the clickstream: Early work on a management information repository at the University of Pennsylvania. In: Information Technology and Libraries, 22(4), 2003, 175-179.
  161. \* \* \*: IBM's Data Mining Technology - White Paper data Management Solutions, April, 1996 (<http://cgi.di.uoa.gr/~rouvas/research/ibm/IBM-datamine.html>)
  162. \* \* \*: NASA Workshop on Issues in the Application of Data Mining to Scientific Data, October 19-21, 1999, University of Alabama in Huntsville, Huntsville, Alabama, [www.cs.uah.edu/~thinke/NASA\\_Mining/DMFinalReport.pdf](http://www.cs.uah.edu/~thinke/NASA_Mining/DMFinalReport.pdf)
  163. \* \* \*: Wikipedia article "Data mining", [http://en.wikipedia.org/wiki/Data Mining](http://en.wikipedia.org/wiki/Data_Mining)
  164. \*\*\* Conferința Noua Economie – O șansă pentru România, organizată de Arthur Andersen și Oracle, sub patronajul Primului Ministru al României, Adrian Năstase, în colaborare cu Ministerul Comunicațiilor și Tehnologiei Informației și Ministerul Industriei și Resurselor, București, Hotel Marriott, 20 martie 2001.
  165. \*\*\* eEurope+ - Action Plan, Europe 2003, A co-operative effort to implement the Information Society in Europe, Action Plan prepared by the Candidate Countries with the assistance of the European Commission, June 2001.
  166. \*\*\* Europe seen leading the way for Internet standard, știre CNN, Web posted at May 17, 2001.
  167. \*\*\* European HLEG report on 'The intangible economy impact and policy issues', after CORDIS focus, 18 december 2000, p.8.
  168. \*\*\* Göteborg Statement Summit of the European Union and the United States of America, 14 June 2001.
  169. \*\*\* Inovation and Technology Transfer, Special Edition, The European Commission, Innovation/SMEs Programme, November 2000, p.21.
  170. \*\*\* MIT to make nearly all course materials available free on the World Wide Web. Unprecedented step challenges 'privatization of knowledge', News Release, April 4, 2001, [http:// web.mit.edu/newsoffice/www](http://web.mit.edu/newsoffice/www).
  171. \*\*\* Nano: the new buzzword in science and technology, CORDIS focus, 20 November 2000, p. 4-5.
  172. \*\*\* Special Issue- The Limits of Semiconductor Technology, Proceedings of the IEEE, march 2001.





## About the authors:

### **Bjørn Arvid Bagge**

BA. Senior Academic Officer. Presently, he works at the Arts and Humanities Library, in the Special Collections of the University of Bergen Library.

[Bjorn.bagge@ub.uib.no](mailto:Bjorn.bagge@ub.uib.no)

### **Pål H. Bakka**

Dr. philos, Senior Academic Librarian and acting Social Science Librarian at the University of Bergen Library.

[Pal.bakka@ub.uib.no](mailto:Pal.bakka@ub.uib.no)

### **Kari Garnes**

Dr. Philos, Library Director at the University of Bergen Library

[Kari.garnes@ub.uib.no](mailto:Kari.garnes@ub.uib.no)

### **Solveig Greve**

Mag.art. History of Religion, MFA (Master of fine arts) photography.

Department head of The Picture Collection at the University of Bergen Library.

[solveig.greve@ub.uib.no](mailto:solveig.greve@ub.uib.no)

### **Richard Jones**

Web and Database Technology Specialist at Imperial College London, he has also worked developing the Bergen Open Research Archive (BORA) at the University of Bergen, and the Edinburgh Research Archive (ERA) at the University of Edinburgh as well as related repository projects. He is also co-author of the book “The Institutional Repository”.

[richard.d.jones@imperial.ac.uk](mailto:richard.d.jones@imperial.ac.uk)

### **Halvor Kongshavn**

Cand jur, Master criminology. He is the Law Librarian at the University of Bergen Library

[halvor.kongshavn@ub.uib.no](mailto:halvor.kongshavn@ub.uib.no)

### **Rune Kyrkjebo**

Dr. art. Senior Academic Librarian, working at the Arts and Humanities Library, and in the Special Collections at the University of Bergen Library.

[Rune.kyrkjebo@ub.uib.no](mailto:Rune.kyrkjebo@ub.uib.no)

### **Ane Landøy**

Cand.philol, Master of History. Arts and Humanities Librarian at the University of Bergen Library.

[Ane.landoy@ub.uib.no](mailto:Ane.landoy@ub.uib.no)

**Susanne Mikki**

Dr. Scient. Senior Academic Librarian, working at the Science Library at the University of Bergen Library.

[Susanne.mikki@ub.uib.no](mailto:Susanne.mikki@ub.uib.no)

**Angela Repanovici**

University Professor at *Transilvania* University of Brasov, Faculty of Mechanical Engineering – Department of Fine Mechanics and Mechatronics, and Faculty of Letters – Department of Library and Information Science. At the same time she holds the position of Director of *Transilvania* University Library.

[repa@unitbv.ro](mailto:repa@unitbv.ro)

**Svenn Sivertssen**

Cand. Scient./Master of Science. Medical Librarian at the University of Bergen Library.

[Svenn.sivertssen@ub.uib.no](mailto:Svenn.sivertssen@ub.uib.no)

**Therese Skagen**

Cand. Polit./Master of Psychology. Academic librarian, working at the Psychology Library at the University of Bergen Library.

[Therese.skagen@ub.uib.no](mailto:Therese.skagen@ub.uib.no)

**Elin Stangeland**

Master of Information Systems. Librarian at The University of Bergen Library.

[Elin.stangeland@ub.uib.no](mailto:Elin.stangeland@ub.uib.no)

**Anne Sissel Vedvik Tønning**

Cand.polit./Master of Pedagogy. Psychology Librarian at the University of Bergen Library.

[Anne.tønning@ub.uib.no](mailto:Anne.tønning@ub.uib.no)

**Maria Carme Torras**

Dr. Philos. Senior Academic Librarian, working at the Arts and Humanities Library at the University of Bergen Library.

[Maria-carme.torras@ub.uib.no](mailto:Maria-carme.torras@ub.uib.no)

**Anne B. Åsmul**

BA. Senior Librarian, working at the Arts and Humanities Library at the University of Bergen Library.

[Anne.asmul@ub.uib.no](mailto:Anne.asmul@ub.uib.no)