Digital Libraries

edited by
Fabrice Papy

This page intentionally left blank

Digital Libraries

This page intentionally left blank

Digital Libraries

Edited by
Fabrice Papy

iSTE

WILEY

FSC
Mixed Sources
Product group from well-managed
forests and other controlled sources
Cert no. SGS-COC-2953
www.fsc.org
© 1996 Forest Stewardship Council

Table of Contents

Preface

Are virtual or digital libraries a step forward in the evolution of those "warehouses of books" which, due to their existence, mission and organization, go hand in hand with the creation and progress of knowledge?

Has the tidal wave of IT only a minor, or in other words technical, impact on the complex structure of a library or does it redefine the libraries' role and mission?

Does the improved access to information based on the Internet and search engines that may be associated with a phenomenon known as "googlemania" overthrow the intellectual and social institution of the ancient world of libraries?

Are digital libraries not the next step towards the future of the information society that moves away from traditional material and looks for new means of storing, preserving, organizing, restoring and, last but not least, visualizing examples of its very own development?

This preface could, without any doubt, fill several pages with simple questions or even sensitive issues on digital libraries. The questions above should therefore only give an insight into the problematic of the impact digital technologies have on the institutions that preserve and organize knowledge.

It is rather difficult to ignore the digital revolution as it is based on the Internet, a very powerful player also widely known as the World Wide Web.

Supporters of the Internet see it as a source for the production and distribution of information as well as a never-ending matrix which offers solutions when improving search techniques and the organization and structuring of data.

However, the network of a digital library based on the democratization of access to electronic resources has rapidly led to changes in people's attitude towards it.

Doubts about the validity of information have been expressed. Librarians therefore need to develop the skill to retrieve valid data. According to Christian Jacob, the search for digital data delivers a number of works which are interlinked. Their complex connection is based on logical presupposition, genealogy, complementarities and mutual explanations[1].

As opposed to the Internet, a library is not only a collection of books which are subject to change, but it has several different missions. One of them is the conservation of books, i.e. archiving. Libraries also represent a privileged environment for intellectual work and research. Users increase their knowledge and adapt their way of thinking to different concepts, as they first need to familiarize themselves with the way the library functions, i.e. its organization as well as the rules and conditions for the use of the library.

This book on digital libraries focuses on new challenges effecting digitized university libraries, public documentation services and privileged partners in teaching and research at French universities.

In the following chapters librarians, information officers, editors and researchers will help the reader to gain an insight into the human, social, organizational, intellectual, political, scientific and technical dimensions of digital libraries.

Fabrice PAPY, Associate Professor, University of Paris 8,
Gil-François EUVRARD, Head Archivist, University of Paris 8.

1 C. Jacob, Rassembler la mémoire. Réflexions sur l'histoire des bibliothèques, Diogène, no. 196, PUF, October–December, p. 56, 2001.

Chapter 1

The Growth of the Role of Librarians and Information Officers in Digital Libraries

1.1. Changes in the world of documentation

Due to the technological and digital revolution, the role of librarians and information officers has undergone important changes.

The transformation of documentation can be established on two different levels:

– for a long time now documents have been produced in a digital format. For 10 years publishers have been using printers that print electronic files which are transferred onto a photocomposition. The programming languages SGML, and later XML, enabled the production chain to work as a digital information channel as well as being able to print scientific documents. No matter whether documents are created in businesses or scientific and cultural institutions, they are all produced in an electronic format. Paper is no longer the only possible way to publish a text;

– the exponential increase of digital documentation has changed people's perception of digital documents over the past two or three years. Digital documents are no longer an exception, they are becoming increasingly common, not only due to the number of them that is created, but also because they can be accessed easily.

Furthermore, the increasing presence of telecommunications and computers (hardware and software) amongst the general public, and especially within libraries, needs to be taken into account. Without this increased access to virtual

---

Chapter written by Christian LUPOVICI.

infrastructures the digital document could never have become so important in terms of access to information.

It is impossible to entirely understand the digital revolution without considering the normalization process that is taking place in all possible technical fields. Opposed to traditional, parallel running normalization processes that have been observed in other economic sectors, this type of normalization process is not entirely transversal but is at least similar for related forms of economic activity, i.e. the way documents are structured or researched on the Internet. Editors, distribution channels, libraries and archives generally use the same, or very similar, systems.

Furthermore, the possibility of retrieving documents which previously were only accessible within documentation centers or libraries has undergone a transformation process. These documents are now made available to a larger audience as they can be accessed from anywhere in the world. Libraries and documentation centers are benefiting from this change.

The Internet has imposed itself as the universal tool used for retrieving information that can be presented on a screen. Its ergonomics are simple and there is no link between search functions and the actual content that is being researched. This type of ergonomics is similar to any other classification system. Experts in the field of information and research state that the simplicity of this type of search engine stops the user from finding the required information. These, however, are very rare cases. This type of search engine has certain advantages over traditional classification systems. Libraries now use the same type of search engine for their own catalogs.

This new kind of search engine has led to a great increase in the performance of access speed and different modes of research. Progress in this field has not yet reached its full potential and important developments are still expected. Currently, these search engines allow information to be retrieved in fields which traditional classification systems simply do not cater for. Specialization still lies far off in the future. At the end of the 1970s and in the 1980s specialists were required not only to work Boolean operators, but also to understand the programming language which was specialized for each database. The use of symbols such as BA = for search entries on the Questel database still lies far off in the future, just as is the case for a "common communication language".

1.1.1. *Transformations in society*

It is not only technology that has changed, but also the general public.

It is clear that we are faced with a different kind of general public which is marked by three main characteristics:

– readers no longer concentrate on one single type of activity. At their workstation they easily move from evaluation tasks to social activities. Work and social activities are highly interlinked. Technological innovations open up these possibilities and the user's behavior is adapting to them;

– the general public likes this type of search as no further requirements are needed. The Internet is used for empirical research as well as for entertainment-based activities;

– this new generation of users is also able to choose not to rely on information systems. This is not an entirely new trend but now technology also supports it. Professionals also have to review their position instead of resisting or ignoring this new trend. The general public actually resolves most questions on their own. However, after working with intuitive search tools the public will become aware of their limits and seek professional support. The reader no longer depends on professionals to retrieve information which could be classed as a revolutionary step forward in the field of research.

Moreover, changes which are introduced by new movements within the field of documentation also have to be taken into account.

1.2. Transformations in the economic situation of libraries

1.2.1. *Too many hits?! The new trend of vague search entries*

Traditionally, librarians and information officers used search engines that retrieved documents which were 100% related to the user's question. The following fundamental problems occur with this type of search engine:

– a full understanding of the user's question is a prerequisite;

– librarians have always found it very difficult to understand exactly what the user was looking for. However, is the idea of completely understanding the user's question not simply an illusion? What if the user simply has difficulties expressing his/her question in a standardized way? A mediator between the user and the system was needed in the 1980s as technological limitations imposed a rather complex language used in the research process. This mediation is no longer needed as the

reader carries out his/her own research by establishing their own search criteria to obtain the required information;

– by finding the exact documents, there is no space left for those hits that might only be partly relevant to the given question.

The information officer's reluctance towards material that is not entirely relevant or the problem of vague research is, however, only happening inside his/her head. How many search engines have been criticized without them even being tested? OLCL was used in the 1980s and was based on two different search keys, author-title 4,4 indicating the first four letters of the author and the title, and 3,2,2,1 which indicated the number of characters of every word in the title. These systems are, of course, old-fashioned but without any doubt highly efficient. They also never produced too many hits. People who are used to online search engines also know how to choose the right key words that will give them the required material in the first three entries on the page.

A system that works very precisely and only delivers results that are 100% relevant to the user's questions requires a professional librarian or information officer. However, general users are therefore unable to use such a system on their own.

Advantages of intuitive search

Since systems are able to limit the number of results, the phenomenon of too many hits within a search process has been regarded as negative for such a long time that its advantages have been forgotten. Intuitive search engines provide the user, who has previously established the boundaries of his/her research, with an overview of all possible data to be found on this topic. The retrieved data includes more information than the user actually needs and not all of it might relate 100% to the given subject. However, all of this data might, to some extent, be relevant to the user's search entry. This is why it provides the user with an overview of all resources that might relate to the subject.

Intuitive searches help to establish the boundaries of the question on their own and indicate ways to ask a certain question.

1.2.2. *The integration of heterogenous services*

The Internet and its easy navigation allows for applications to become more interdisciplinary. The user moves from one application to the next and sometimes also moves towards more entertainment based activities. This is why the OPAC

(Online Public Access Catalogue) needs to be created as some sort of portal that allows access to many different applications.

1.2.3. *The librarian's challenge to reach customer satisfaction*

That a user might be happy with the responses obtained from the search engine is something rather shocking and worrying for librarians.

It is shocking because librarians know that the user of the system has missed out on relevant answers, not only from the system used because he/she lacks experience in research, but also because other systems of documentation that are different from the one used have not been taken into consideration.

This is worrying because the user is satisfied with incomplete information, as the research target was badly set. The user even accepts the risk of missing out on important aspects.

Furthermore, librarians see that the user ignores, or remains indifferent, to the role they play within the library.

There are certain points to be made with regards to this phenomenon:

– the average user looks for one answer, not all possible answers. The librarian's obsession with delivering all possible answers is a surplus in quality that is not appreciated by the user;

– librarians are only ignored when the research to be carried out does not cause any problems or is carried out on data that is not complex. This is only the first step in research. Librarians can therefore concentrate on complex problems where their help is needed and where their intervention makes sense;

– librarians are not completely deprived of the research process, but no longer play the role of the mediator between the user and the system. Librarians fulfill their role when structuring primary data and metadata which is classified by search engines. Their tasks are focused on the more sophisticated levels of the process of documentation.

1.3. Changing a paradigm: changing the object "information"

1.3.1. *Breaking with the traditional way of managing physical objects*

The transition from an analog to a digital environment can be considered as a cultural revolution since the physical aspect of objects is changing.

1.3.1.1. *From a manufactured object to information objects*

Traditionally, libraries consist of manufactured objects such as books, magazines and CDs that are classed as volumes, series or disks. Their content is subject to the support which is used to store the information.

In the digital environment, the document is independent of its support and leaves room for experiments. The interpretation of its content, its structure and its additional metadata gives it its own force in the field of documentation.

1.3.1.2. *From an exogenous catalog to endogenous metadata*

Traditional documents such as paper, tape and disks do not comprise elements of documentation that enable the user to retrieve them. This is why librarians and information officers have invented certain techniques and rules to classify and retrieve these documents. The elements used to classify documents and create an index are, of course, taken from the documents. However, this is done in an artificial way by a librarian who will create data and transmit it to an exogenous catalog.

From the minute digital documents are created, they contain descriptive elements that enable the user to classify and manage them. This metadata is endogenous, i.e. integrated into the document. Furthermore, within the process of documentation additional metadata is integrated into the document. The general trend is moving towards a completely automatic process of administrating data, as the capacities are far too large for any human intervention. The way this data is processed therefore needs to be programmed. Here again, the librarian's expertise is required to process the given data.

1.3.2. *New objects in documentation*

Dematerialized documents that do not require any form of support

The first step in the digital revolution is the dematerialization of a document. This does not only mean that the document can be accessed from any possible location, which introduces the concept of ubiquity, but also that the processing of this document is carried out automatically. It takes a tangible form which allows for the visualization of its form, its reality and its integrity.

Furthermore, as the document does not need to rely on any kind of support, the importance of the layout changes. Its layout is subject to its structure.

Without a stable support, the question of how to materialize the document is not a question of how stable a support is, but only deals with the question of which role the content plays within its environment.

Hybrid objects (multitypes) represent the next generation of objects. They are based on multimedia and can read texts, icons, maps, videos and even 3D images as well as tables and programs. The term multimedia, which was previously used in a different sense, finally measures up to its meaning here. Complex objects are not only created, but are also managed and presented to the user.

These new documents are structured documents. The way the information is structured, based on XML's standards for example, integrates the semantic significance of the document. This enables the user to retrieve information inside the document, i.e. internal navigation.

The adjustable level of granularity in these objects is also a revolutionary trend in the field of documentation. Traditionally, libraries or documentation centers only addressed one level of granularity. These are the number or title of a series, an article in a periodical or the chapter of a book. Traditionally, libraries or documentation centers only processed one level of granularity. A book and a reference, an article in a magazine, and a video as well as an illustration and a manuscript, are now processed simultaneously. This form of processing is based on the use of tree-diagrams.

Direct access to pieces of information is a great advantage of this concept as it is also possible to preserve the visibility of the entire document and therefore keep its coherence.

Understanding all functionalities and transformations in concepts and techniques is vital for librarians if they would like to gain full insight into the transformation of tomorrow's documents on the Internet, which works very differently from the librarian's traditional research techniques.

1.4. Changing a paradigm: information in a network of documentation

1.4.1. *Information is linked to a network of information*

Digital documents can no longer be a simple back-up of paper documents. They should be navigated with the help of internal links as well as links that are created between different documents. This technique goes further than the reference system used in paper documents. It is a philosophy that includes alternative routes and allows for the personalization of the order in which the information is read depending on reactions of the public (e.g. open lectures) or a previously established set-up.

With the digital document being "active" it can actually process itself. Linguistic structures or statistics, in particular within the semantic environment of the document, could be automatically linked to the document therefore allowing for a better understanding of the document.

1.4.2. *Processing a high flux of dematerialized information*

The fact that traditional administration systems in libraries are no longer used also stems from the large amount of digital documents that are transmitted to libraries. It is therefore impossible to process every single one of them, even more so since the documents only consist of files and bits.

Automatic processing is therefore required to enter data, process it and manage it. Visitors to a library have to use programs to deal with this enormous amount of information.

Professionals intervene at a higher level. They set the parameters for the automatic manipulation of this information.

This also means that the trend is moving from manually administrating every single document towards an era where sampling is added to manual administration.

Digital processing of documents requires the introduction of identification procedures and troubleshooting (faults, rejection, etc.). These are of a rather industrialized nature.

1.5. A new way of organizing libraries: the impact of the digital revolution

1.5.1. *Impact on the functioning of a library*

If the traditional aims of a library are to remain the same, research techniques have to undergo profound change due to the developments that are linked to the appearance of digital documents.

The task of acquisition can technically be subdivided into two different forms:

– harvesting and the appropriation of the entire content, or parts of it;

– the acquisition of files requires an at least technical agreement between the supplier and the client.

Processing data will then include the following steps: extracting metadata from the documents received from outside the institution, as well as from documents

produced inside the institution, structuring these documents, and tagging and enriching the metadata.

The storage capacity of these systems is linked to the architecture of the information system used. It is, however, extremely important that librarians are familiar with this system. They therefore need to broaden their knowledge and skills in this field.

The library's task of preserving information cannot be confused with the pure storage of it. The preservation of data, even in the digital era, remains a task that can only be carried out by a library. Documents therefore need to be processed in a way that enables the user to find and exploit them in the future.

The job profile of librarians and information officers therefore adapts itself very well to researching digital data. Research strategies, such as hyperlinks, need to be reviewed by professionals in the field of documentation, including the usage of the most powerful technology that is available today. This new trend needs to be taken into account and the readers need to be supported during this transition period.

1.5.2. *Impact on the concept of information*

Digital documents work "intelligently" if they are programmed in an intelligent way. This is why the acquisition of know-how is needed in the fields of tags, integrated metadata, weighting system, and how widely the document is known in the scientific community. The concept and the perception of information are changing.

1.5.3. *Impact on distribution*

The distribution of a digital document on the Internet can no longer be compared to the traditional distribution of paper volumes, even if they were categorized and accessible from several libraries. Its distribution has increased 1,000 times over. This increase can be applied universally. This is why digital documents raise so many questions that could already have been applied to paper documents but were never discussed to such an extent because their distribution remained more limited.

1.5.4. *Impact on intellectual property*

The question of intellectual property, copyright, quotations, etc. has a negative impact on the researcher and editor's image of electronic documents. Even if these problems increase due to the high speed and universal access to electronic

documents, digitization also allows for those problems to be addressed far more efficiently. If the document is not public its access can be protected. Above all, it is possible to link access authorizations given on one single document to general access authorizations that are valid for the general public. A document can therefore only be accessed by its target audience. Abuse is quickly established by the server. In the case of public documents, it is important that they are used by other people, while at the same time the author needs to be protected from plagiarism. The idea that plagiarism is much easier with documents that can be accessed on the Internet is not a valid one since the intellectual property of a document that is quoted, read and therefore widely known is much easier to protect. Furthermore, it enables the researcher to give a perfect quotation by using the copy/paste function and therefore minimizes the risk of giving a false quotation.

1.6. New trends

1.6.1. *Introducing administrative aspects of documentation into the document*

Digital documents are far more than just an edited document. They are also objects of documentation, i.e. pieces of information which can be accessed and inside which we can navigate to its metadata. The way in which the document is structured is significant; its tagging and the introduction of administrative metadata must be carried out with great care when the document is created or handled later on.

The document is not an isolated entity. It is part of a network of a series of documents accessible on the Internet which is now transforming into a documentary system. Documents which are published on the Internet are connected through links in the hypertext. In the case of archived material digitization, this often leads to problems in identifying a specific document amongst all others. In some cases a website has to be considered as a tool in the documentation process similar to a collection of articles.

1.6.2. *The librarian's role in the editing process*

The librarian's know-how is increasingly required in the editing process. In the future, librarians will assist authors, especially in the field of university-based research, in order to help them to produce intelligent documents. They will also work with editors and underline the fact that additional documentation is of great importance to a document, even if it is fiction (e.g. bonus tracks such as documentaries that give added value to a DVD).

Previously, librarians played an important role in the institutionalized diffusion of information. Their role is changing and they will also be involved with, and guarantee the quality of, information distributed on the Internet.

The most important sector requiring librarians in the future is preserving digital documents in the long-term. In this sector the tradition of libraries is needed to develop the necessary know-how, as a library gives priority to cultural aspects over economic interests.

Future job descriptions for librarians and information officers will clearly contain the fact that they represent the link between the library and the publisher.

1.7. The digital library

1.7.1. *The virtual library*

A virtual library is a library where documents are virtual, i.e. without a stable support. There are several advantages to virtual libraries.

First of all, they are dematerialized and their location therefore becomes insignificant. Digital libraries have the advantage of universal access.

Digital documents can also be accessed by several users at a time. Thousands of people can visualize and download a document without hindering others doing the same.

This type of library is far from being homogenous. Its sources, structures and documents are highly diversified, they stem from internal (institutional) production, commercial production, online collections etc.

Their documents are real multimedia documents, i.e. hybrids of different forms of media. These documents need to be organized, and standardized access to them must be provided.

1.7.2. *A "real" library*

A digital library is a real library since its collection is organized, selected and well presented. Its documents are processed and administrated and their access can be controlled. A digital library also responds to the need to develop collections of documents such as digital resources, articles, books, etc.

1.8. Introducing different layers to the core sector of the profession

Rather than believing that the profession of librarians will change, which is only applicable to some cases, the introduction of different layers in the core sector of the profession should be taken into consideration. The core sector of this profession re-invents itself each time a new field of work is discovered.

1.8.1. *Support for online library users*

The librarian's task of advising and offering support to users is also changing within the digital environment. Librarians work in shifts throughout the world to provide a service which is up and running 24 hours a day, seven days a week.

Legal questions regarding the creation and distribution of documents, which were already worrying in the times of traditional libraries, have come a step further in the digital environment as it increases distribution.

1.8.2. *Providing training for users*

If providing training for users is part of a librarian's duties, this task becomes much more diversified with documents developing an internal structure.

Other than knowing how to use traditional resources, users also need to be shown how to use electronic sources. The latter does not replace the use of traditional resources, but completes them.

1.8.3. *Managing materialized objects as well as digital documents*

Managing materialized objects is a principal part of a librarian's work. However, they also need to know how to manage virtual documents of a complex structure and how to work under a certain format. Librarians have to become experts. Limits have to be established that indicate when analog systems are an advantage and when they become a handicap. It is up to the librarians to conquer this new territory.

1.9. Broadening skills and responsibilities for all of the library's staff

1.9.1. *Managing old and new techniques simultaneously*

As previously illustrated, new technologies do not replace the old ones, they actually complete each other. In libraries the traditional and the new technologies amplify each other.

Librarians need to pay attention and react to new techniques and methods that come up in their field of work. On the other hand, they do not have to forget about their traditional tasks as the documents they are in charge of, as well as their clients, belong both to the old and the new era.

By learning how to use traditional methods and understanding the cultural aspects of a library, librarians will find it easier to use future technology without jeopardizing the library's essence.

1.9.2. *Increasing qualifications and responsibilities*

The digital environment requires an increase in skills as these new competencies need to be added to the old ones. Furthermore, an increase in quality is needed as a digital working environment requires a higher level of analysis.

In the digital environment the librarian's responsibility in terms of content increases since their choices are made public when the documents are distributed on a large scale.

The digital environment increases the relevance of the institution's strategic and political decisions.

With new technology, these choices become increasingly risky as the position of the institution partly depends on the library's capacity to transform itself into a modern institution.

The legal responsibility of every single librarian, and not only their director, is based on the increasing number of documents produced by the institution that are distributed online. Access to these documents also needs to be protected and access conditions must be created.

During this transition period librarians need pedagogical capacities when creating digital documents, i.e. the technical side of production as well as a profound knowledge in the field of heuristic research in the digital environment.

The digital environment requires a higher level of technical skill not only to carry out editorial tasks or those linked to the organization of a library, but also to make the right strategic choices for the library as an institution.

Chapter 2

The Tao of the Digital Library: A Library Without a Librarian?

"Thirty spokes share the wheel's hub. It is the center hole that makes it useful" [Tao-te-Ching, Book I, 11].

The image of a wheel that turns around an empty space could represent the digital library of the future – a virtual space, a portal that allows access without mediation, a platform of information, a tool for users who are far away from the actual library. It is an efficient tool, a wheel that turns a vehicle which moves. But what is to be found at its center?

"It is the center hole that makes it useful." The conclusion of Laozi's *The Book of the Way and its Virtue* could be the ancestor of marketing strategies published by STM editors and suppliers of content management systems. What is their main argument when promoting their products? Their latest products and services are directed towards the end-user, i.e. researchers in front of their computers. In their eyes the actual space of a library and the work of librarians is rather unnecessary and will disappear sooner or later; to them it is a niche in the market.

The center hole, i.e. an empty space, is currently being created. Libraries will first be transformed into a virtual space and later into learning centers, i.e. multimedia resource centers. They will become a feature of education and teaching, just as e-learning is an accessory to traditional teaching methods [BEN 04]. With the

Chapter written by Joachim SCHOPFEL and Jacques CREUSOT.

"hyperelectronification" of campuses [ARN 03], information has now moved from inside the library to the outside world.

The librarian is called an information officer, or even facilitator, and finally becomes a technician specialized in the field of troubleshooting or carries out rather simple tasks at the helpdesk. Following this logic of relocating traditional tasks, the next step would be to replace libraries with call centers located in Morocco or India!

Debates at conferences often call into question the librarian's professional identity. Where are we going? Do we know where to go? Do we want to go there? Is it only a simple image problem as Bernat [BER 03] suggests? Is the disappearance of the library as we know it not simply a very negative and unrealistic outlook on the future?

"Is there a future for librarians now?" In 1990 Dick Fletcher (New Media) asked this question on the occasion of the 13th conference of the UK Serials Group and even now it is a relevant question since it has still not been answered.

At the beginning of 2004 Bruce Heterick from JSTOR ended his report on the future of libraries with the ironic statement: "There are three types of people in the world: those who make things happen, those who watch things happen, and those who don't know what hit them." [HET 04].

Let us try and understand what he meant with "those who don't know what hit them". Let us take a look at some aspects of the transformation that the job profile of librarians and other library staff has undergone. Where does the concept of a digitized library come from and what is the driving force behind it? What kind of impact has this revolution had on the library's professions and the skills of those professionals? How can this revolution be accommodated by activities within the field? Which direction should we move in? What should we do if we do not only want to "watch things happen"?

2.1. The technological supremacy of the concept of the "digital library"

The concept of the digital library appeared in the middle of the 1970s but only gained increasing popularity 15 years later, i.e. at the end of the 1980s. At the beginning, this concept was entirely based on IT and had nothing in common with the library as a work environment where people study and carry out research in order to broaden their knowledge.

After analyzing the publications of scientific journals since 1990 (a sample of 1,086 articles that appeared in 13,000 journals was used) the following figures were

established. 75% of all articles published in digital libraries appeared in IT journals, 15% in journals that focused on humanities and social sciences, and only 5% in journals specialized in documentation and the organization of libraries.

When checking the year of publication it becomes evident to what extent technology dominates the concept of digital libraries. 75% of all articles on digital libraries published in the field of human science appeared in the last five years, which shows a large gap between human science and the field of technology. Librarians are even further behind when it comes to addressing the subject. Most studies on the organization and human resources of digital libraries (e.g. positions, training, recruitment) only appeared in 2002 and 2003.

When the human factor becomes the object of scientific studies this implies the following questions: to what extent does the user accept new services and functions and how does the interaction between humans and a machine work (system, ergonomics)? Studies that focus on professional librarians as the main players and not simply as a technical operators are recent and relatively rare.

At the beginning of digital libraries there was technology, engineering, as well as the development and the installation of tools. Professional librarians who are directly concerned by this new technology ignored this development for too long. Studies on how, where and by whom these technologies could be used, as well as research on the impact on human resource management, only appeared 5 to 10 years after those carried out on computer science. On this point, the digital library had already become a reality in the field of IT. At the occasion of the 3rd congress of French information officers, *Le Monde* stated that "technological developments were a step ahead of attempts to solve problems between humans" [DEG 79]. This asymmetric development has had an impact on the library professions up until now.

With the birth of personal computing and the introduction of TSI in Germany, big university servers were declared to be part of the "computer scientist's tasks as they are part of IT". Information officers and librarians suddenly found themselves in the role of the end-user which inversed the ratio between supply and demand. This trend is similar to what happened in other industrial sectors where technical progress determined behavior and therefore new requirements that also had to be satisfied.

On the one hand, IT sets out a framework that determines the structure for the content and the tools to be used. On the other hand, people use these technologies to create and distribute a certain content or access it (peer-to-peer). The first digital archives were set up by researchers and computer scientists, not by librarians. This sometimes leads to the disagreeable impression that librarians are only used as beta testers of new IT systems.

2.2. TSI's influence on the market

The recent increase in the speed required to access information online is not the only driving force behind digital libraries. In the context of a financial and economic crisis, another factor has an increasing impact on the development of libraries: it is a commercial strategy of companies that create and publish scientific information. After heavily investing in the field of research they seem to have decided that mediation by other professionals such as agencies or libraries is no longer required.

It was not that long ago that people had to enter a library in order to access resources. Searching a database required specific skills catered for by professionals working in the field of documentation.

Those times have changed. After studying behavior and requirements the producers of TSI developed services that directly address the end-user. The functions of these services are simple and intuitive, a Google effect can be observed. With the emergence of a search function that combines research and access to information in one single function, every other type of search engine ends up with a marketing problem [ARN 04]. Search engines like Google and Scirus are integrated into a portal or platform. This link enables everybody to find the required information without having to consult a professional.

From a professional point of view, this form of research does not necessarily guarantee the quality of the sources given, neither their relevance nor how thoroughly they address the respective topic. Baltz, for example, states that information officers should not be frightened by the "scary amount of autodocumentation" since there will always be a place for them in between IT and the communication of ideas [BAL 03]. From a user's point of view, this argument remains theoretical because, due to the ever-increasing performance of research tools, this type of autodocumentation corresponds to his/her needs. The success of a portal with a simple appearance, such as VASCODA which provides access to multiple German TSI resources, seems to back up the user's point of view which is mentioned above.

Currently TSI's producers still need professionals that sell their products. The supplier-client relationship is currently changing. Large TSI companies (Wolters Kluwer, Reed, Thomson, etc.) see themselves as education companies. Their client is not the library but the scientific community. Alternative editors such as BiomedCentral or IOP are directly financed by researchers/authors or their institution without having to rely on the acquisition budget of libraries. Integrating the costs of TSI into the budget of research organizations is a leitmotif of the open access movement. Only very recently have STM editors also taken an interest in this

new concept. What is happing to the library as a center that purchases new publications?

When an editor like Elsevier organizes a seminar for librarians, content and services are no longer discussed. Instead, new functions are presented and explanations are given on how to promote these products in the respective community. In this context, a "good librarian" is someone who manages to increase the use of these resources and secure the loyalty of the client. The library is transforming into a commercial platform of TSI producers. The competition of the "best advertising" which was introduced by an important STM publisher is an example of the negative aspects of this transformation.

2.3. The virtualization of a document's function

Traditional scientific and technical information has only one aim when integrating information technology and telecommunications: to improve, research, exploit and evaluate the information received and transmitted by the researcher. The entire problematic of the function of documentation can be resumed in the two prepositions by and for. Is there a link between the supplier and the end-user of information? If yes, what kind of link is this?

Without a doubt some scientific circles need databases, catalogs and a personalized service. But who manages the technical features of common access, technical support and the development of protocols? Who decides upon and co-ordinates the acquisition of resources? Remote access and developing efficient tools (e.g. search engines that search several databases and mulitprotocols via portals) shift the focus onto the user's needs when it comes to documentation services. However, who controls access, who decides who uses what and how?

Shifting the focus to the user's needs is generally not a bad choice for libraries, but when analyzing the situation in great detail, it becomes evident that if service providers are responsible for creating libraries, these libraries only consist of marketing strategies. If the needs of scientific circles are addressed by the suppliers of information, the library transforms into an additional service of the TSI industry without having any added value to it.

According to the last opinion poll carried out by JSTOR covering 7,000 teaching fellows and researchers in the field of human science [HET 02, HET 04], the actual location of the library was the last point mentioned when being asked about carrying out research. It was far behind search engines and other specialized services (portals, databases, etc.).

American researchers and teachers consider the function of the library as still being important, especially when it comes to buying resources (budget), but less so when it comes to access (gateway) and archiving. However, all of them stated that their research was not very dependent on the library, especially if they were not working on large campuses. Generally, they estimated that the importance of a library will decrease in the next five years. "In fact, many faculties can foresee a future in which they will never actually go to the library" [HET 02].

Researchers and teaching staff often only give the library a polite nod when it comes to researching information [ARN 04]. What is important to them is access to electronic resources and the guarantee of durability for that access. They require an archive. However, in the electronic era this archive does not necessarily need to be in the library or on the university's campus. In the words of Tao-te-Ching, it could also be in the center hole.

2.4. Development and changes to job profiles in the CNRS directory 1982–2002

The Tao also says not to pity feel sorry for yourself if you want to endure [Book I, 22]. Let us now see how this idea can be applied to the professions of a library. Several approaches have been chosen for a better understanding. The development of professions in the CNRS, France's biggest research organization, and those of the INIST, as well as changes in training programs in the UK will be analyzed.

For more than 20 years the professions and positions within the CNRS have been regularly subject to formalized studies which are published in a directory. When comparing the directories of 1982, 1991, 1998 and 2002 the following question arises. What is the impact of new technologies on libraries and documentation centers?

From 1982 onwards job profiles in the CNRS reflected the way laboratories used information and documentation services. This phenomenon could be observed on all possible levels, from the simple controller who updated a computer-based catalog to a documentation engineer who managed the entire system and adapted it to new technologies. Another job profile could be found in the directory of 1982 indicating a librarian who was "specialized in the automatic organization and management of collections".

In the middle of the 1990s the informatization of the workplace as well as activities and skills became increasingly important. Knowledge of documentary IT and the know-how to create project specifications and choose the required tools became one of the indispensable skills of category A library staff. Only in 2002 did the CNRS's directory become similar to the REFERENS, the directory for higher

education [HIC 04] and indicated that the digital library had become a reality. There are three examples of this trend:

– simple controllers and technicians need to know how to convert printed documents into digitized documents;

– engineers need to know how to create electronic content management systems;

– from the level of engineering assistants onwards (a diploma of higher education being a prerequisite) a good knowledge of the legal environment is required. We will come back to this later.

This adaptation can be subdivided into three different levels:

1) The amount of IT used for different activities is now affecting more job categories. IT that was previously only used by job category A (engineers, assistants) is now also used by categories B and C (technicians, controllers). The knowledge and skills in the usage of IT which were part of the job profiles of engineers in the 1980s were suddenly also part of technicians' job profiles. In other words, around 10 to 15 years later technicians needed to have the same skills and knowledge that was expected of engineers in the 1980s. Furthermore, the level of the required qualification has been increasing as Christian Lupovici observed [LUP 04]. We must add that this process is still not being reflected in the salary structure.

2) New technologies make the old boundaries between documentation and the organization of libraries obsolete. Studies on job profiles within the CNRS from 1995 to 1997 have shown that the fields of technicians and assistant engineers have been merged. The human resource department was not brave enough to actually merge the job profile of engineers and researchers even though studies showed that in reality these two fields were already starting to integrate. In 1998, on the other hand, a new type of position, a hybrid between the fields of documentation and IT, was created – an engineer and administrator of databases in documentation. In 2002 the traditional three fields of information officers, librarians and archivists were losing their importance due to this newly created position.

3) Technology's rapid evolution demands skills that are not part of initial job training. 2002's directory not only contains the introduction of new technologies, but also states that TSI professionals need to "take part in regular training sessions in order to remain up to date with new technologies in their field". In other words, keeping up with technological innovations, which was already mentioned in 1998, and continuous training has become a principal part of their work, as this is vital for the organization as a whole.

This is especially applicable to the legal environment (e.g. authors' rights, copyright, legal obligation when keeping documents, intellectual property) which,

according to 2002's directory, should be part of the skills for positions within category A. When it comes to technological developments and the know-how of professions linked to a library, the library as an actual place, i.e. a physical space, seems an outdated idea that is no longer relevant. Recently, Didier Frochot claimed that the professions of IT documentation were "not very well informed on legal questions" or even "completely ignorant towards them" as they lacked serious and thorough training. He particularly accuses the institutions that provide training for these jobs of only focusing on technical aspects while ignoring organization and management. Moreover, he states that before actual changes in the profession will take place, the librarian's employers should organize the acquisition of this knowledge. ISIDROIT is an example of how this criticism could be taken on board. This interdisciplinary network comprises the CNRS's information officers and is subdivided by region. In the Rhône-Alpes province it is called ISIDORA.

2.5. Supporting professions – the INIST approach

With 350 TSI professionals, 100 of whom work at the library, the INIST (Institute of Scientific and Technical Information of the CNRS) is directly affected by all technological developments [SCH 03]. From 1998 onwards the INIST has produced reports on the challenges of new technology and online information. With the help of a large number of its employees, the INIST 2000 project set out the main standards for the acquisition, management, storage and distribution of electronic resources, as well as transforming its products and services in order for them to be able to enter the Internet era [INI 99].

In order to observe and anticipate the development of job profiles and skills of TSI, the INIST introduced a permanent training program known as "compétences métiers" in 2002. This project is directed by a committee which consists of human resource managers whose chairman is also the director of the INIST. This procedure avoids problems that have been observed elsewhere, i.e. the "absence of any form of management, no defined positions, no idea of future developments nor the existence of acquisition strategies" [BEN 04]. Instead, a decent support for the staff involved with the project should be guaranteed. This approach is carried out in three steps:

− creating a directory of skills and professions the INIST currently requires. This directory is based on that of the CNRS and is enriched by the ADBS's directories (ECIA);

− an impact study on current developments in the field of TSI is carried out on the basis of this directory;

− suggestions which should function as guidelines in decision-making processes, as well as scenarios on the development of skills required by certain professions will be made in the immediate future (3 to 5 years).

Developments in IT and telecommunications, users' needs, products and services of the INIST must be observed. Furthermore, all players need to be interconnected during all steps of the project, i.e. there has to be good internal communication between all parties. This project is being carried out for all professions in the fields of documentation, library, IT and communication.

Job descriptions in the directory are discussed in all of the INIST's departments and are then put together as a coherent document. This writing process goes hand in hand with discussions in the individual departments which will then define possible development scenarios. Every scenario will need to state the skills the department needs to acquire in the future. Once this work is completed and put into one document, it will serve as a decision-making tool as it allows for the visualization of current skills that will help to determine training or recruitment.

This is a long-term project which is closely linked to TSI developments within the INIST and the CNRS. It also increases the quality of libraries and advertises the concept of the digital library. Questioning traditional methods and activities (e.g. cataloging, periodical subscription purchase and renewal) sometimes leads to negative reactions, but on the other hand opens up new perspectives for departments as well as for individuals. It leads to the merging of several activities, changing of procedures and opening up towards the outside world.

The project is also based on an exchange with other public organizations (e.g. INSERM, CEMAGREF, INRA, INRIA, CIRAD, IRD). The aim is to prepare the staff to take on different responsibilities in a new and often complex environment, or give them responsibility for other tasks. The think-tank, in co-operation with the CEMAGREF, focused mainly on the relationship between the researcher in the field of documentation and the management of knowledge.

Every project in a digital library has its own "biography" [GRE 02]. That of the INIST's library follows several directives:

– instead of creating new structures, new activities and skills are integrated into acquisition, management and production services;

– a limited number of new positions is created. These pilot positions are created to co-ordinate the development of activities and skills for all functions (e.g. editors of electronic magazines, researchers responsible for licensing, digitization operators). This development goes hand in hand with the employment of interns from local universities (from 2nd year students up to students studying postgraduate courses). They work on specific issues such as managing electronic magazines, analyzing user statistics and the creation of databases with open access;

– new training requirements (e.g. legal environment, TSI, IT and telecommunications market) are analyzed in the individual departments and

integrated into the annual planning for training. As external training cannot be provided, part of the training is provided internally or is carried out by the employees themselves (e.g. electronic magazines, copyrights, managing licenses, statistics, the production of catalogs);

– the transformation process of the library's function is spearheaded by different projects which are carried out by specialists in the respective fields (e.g. producing catalogs, delivering documents in PDF format, managing access authorizations).

Compared to Greenstein's and Thorin's [GRE 02] "biographies", the Institute's digital library project is still in the start-up phase and therefore of an experimental, competitive, manager-based, dynamic and innovative nature. This project is attempting to find the ultimate solution known as the "killer application" that will allow for a reduction in the complexity of IT systems and all the problems linked to them, as well as being a cheap alternative to traditional methods.

However, this project is currently stabilizing. Today's digital libraries are complex and consist of different modules, such as several archiving systems both internal and external (e.g. CCSD, the CNRS's service) and an architecture that comprises online services which are based on several different applications (e.g. content management systems for libraries, platforms for archiving and distribution, online databases and portals). For librarians it is not always easy to keep up to date with the rapid developments in IT systems and sometimes training is a reminder of the story of the race between the tortoise and the hare – there is no point in running, you just have to leave early enough… The library of the INIST is trying to address this problem by re-defining the heart of the librarian's profession and defend its functions, i.e. analyzing users' needs, selection and acquisition of content, cataloging, management and archiving of resources.

2.6. A new job profile is emerging – the e-serials librarian

At the end of the 1990s a new job profile emerged in the libraries of anglophone universities known as an e-serials librarian or an electronic resources librarian. Often the creation of these new jobs went hand in hand with the merging of teams responsible for new resources with teams managing existing resources.

After analyzing the life cycle of an electronic journal, the INIST's library team responsible for this project determined the boundaries of this new role. These boundaries are rather wide. They include the development of an e-strategy for the library, evaluation of the editorial offer, negotiation of licenses, training of colleagues and users, management of links and access authorizations, technical assistance and troubleshooting, analyzing user statistics and the promotion of new services.

Some functions were specified and subdivided into special tasks. These are market research of TSI online (e.g. keeping up with technological innovations, observing prices and economic models), analysis of client needs, i.e. needs of clients of the INIST and the scientific community of the CNRS, negotiating and analyzing contracts, creating a licensing system and evaluation scales for reading, managing the relationship with editors and agencies, ordering documents with and without a mediator, saving and administrating access and controlling usage with the acknowledgement of recommendations, i.e. COUNTER [BOU 05].

Some functions are taken over by other structures within the INIST, such as the IT department (e.g. access management via a proxy server), a new department in charge of portals (e.g. creating portals, analyzing problems, information on access authorizations) or the technical support unit. In order to make things work, responsibilities have to be negotiated and good co-ordination and communication between the individual services is required.

At the same time, the INIST's council of internal documentation, the acquisition committee, has started to integrate electronic resources into their work and discussions in order to develop a coherent strategy when it comes to the acquisition of electronic as well as paper-based resources.

This work on the profile of e-serials librarians has led to the creation of several new positions in the library and will have an impact on the way services are presented. However, this job profile introduces certain problems:

– the technical level: how much knowledge in IT is required? This question becomes especially relevant when web applications such as virtual libraries are created, or electronic resources are integrated into information systems. How can the power struggle between IT and the library be won by the library when it comes to analyzing requirements, defining tasks within a project or defining functionalities? Do we need "second class" IT consultants whose training is inferior to their colleagues but who still have a good knowledge of technology? The interface between the user and the library also remains to be invented;

– legal knowledge and skills: is the librarian in charge of licensing still a librarian or are we moving towards a job profile based on administrative or legal tasks, with no degree in this field?;

– relationship with the reader: if the library is transformed into a portal, what is the link between the librarian and the reader/user? Who manages and filters communication? How can this be organized? Last, but not least, who actually is the user? Is it university researchers and teaching staff, individual users that define their status via licenses, a network of information officers in their laboratories or scientific administration?;

– the level of decision-making: who summarizes the analysis of requirements and decides which negotiations are carried out? Financial interests and a national co-ordination of acquisition limit the decision-making power of librarians as 50% of the budget is spent on acquisition, i.e. the decision is made on a political level far from the actual library.

These questions are neither theoretical nor conceptual; they have a practical impact on the career of library staff (e.g. entrance exams, development within the individual's career, mobility).

In the long run, training, experience and skills will provide answers to these questions. Answers will also be provided by professional librarians working at the INIST's library, by TSI strategies of the CNRS and by a national body that will comprise all universities and research bodies and provide access to electronic information.

2.7. Developments in training requirements – the UKSG workshops 1990–2004

Another way of observing developments in job profiles is through studying training and education in the field. With the emergence of the Internet in the 1990s new requirements within training also appeared. In France the ADBS (Association des professionnels de l'information et de la documentation), a professional association for information officers and library staff, provided new materials and a greater diversity in the training offered by co-operating with INTD-CNAM [GIC 03] from 1994 onwards.

In this chapter a very different body will be analyzed – the UK Serials Group (UKSG). For 25 years it has united librarians, editors and intermediaries and technology vendors in the UK and in other countries. Once a year the UKSG organizes a conference that comprises up to 600 professionals from a range of different fields. The papers, debates, product reviews and workshops from these conferences give an excellent overview of current questions and projects. We have studied the content of 183 workshops from the conferences that took place between 1990 and 2004 in order to obtain a better understanding of the training requirements in the field of digital libraries [SCH 04]. Some of the results that were established during the analysis are as follows.

First observation: the number of workshops and sessions has continually increased. At the beginning of the 1990s five to ten workshops took place at each conference. During the latest conferences 15 to 20 workshops with up to three sessions took place. This not only proves the success of this type of exchange and

training, but also reflects the increasing need to obtain further training and to stay informed about recent developments.

When comparing the content of these workshops to the functions of JSTOR-based research without making a difference between paper and electronic resources, significant developments can be observed. The number of workshops on the library's role as a buyer has decreased from 50% down to 15% while their function as a gateway becomes increasingly important. The number of workshops on this topic has increased from 25% to 65% of all workshops. The third function of the library being involved in archiving only came into play from 1998 onwards. 5% of the workshops during the last three conferences dealt with this topic.

While the digital library only played a marginal role in the early 1990s, today 90% of all workshops deal with this topic. This proves that digital libraries now play an important role when it comes to further training for librarians.

Let us now analyze this phenomenon in greater depth starting with the way it is linked to "JSTOR functions". 15 years ago only one-third of workshops on acquisition dealt with the acquisition of electronic resources, while currently 70% of all workshops address this subject. Since the beginning of the 1990s the concept of access has been linked to the emergence of the digital library. Since then the percentage of workshops dealing with the library as a gateway has remained stable, ranging between 75% and 85%. Workshops on the function of archives now solely address the subject of electronic archiving (100%).

To end this analysis let us take a look at the content of 100 workshops on the topic of digital libraries. A third of these workshops offered training on programs and equipment such as developments in content management systems (e.g. organization of e-journals, automatization of the organization of periodicals), web applications (e.g. search engines, creation of websites), new logic solvers and, last but not least, digital rights (e.g. digital rights management – also known as DRM, access control). 16 workshops presented the editors and the products they currently have on offer and the market of electronic journals, a few of those workshops dealt in particular with the consortium NESLi. Others worked on electronic books, e-commerce, or on the impact the concept of open access has had on scientific editions. 15 workshops dealt with legal questions such as copyright and licensing.

Ten workshops dealt with developments in the field of catalogs (e.g. rules and standards, OPAC) and the format of documents and metadata (Dublin Core). Seven workshops, such as the COUNTER project, worked on the exploitation of user statistics.

However what about job profiles as a subject on its own? Before 1998 this topic was not covered at all. From 1998 onwards 15 workshops have examined several subjects linked to human resources (e.g. communication, ongoing training, and changing relationships with clients) and scenario planning, i.e. new forms of organization within the library.

When taking a step back, the extent to which training is offered during these conferences shows the broadening of skills and the changes in job profiles Lupovici [LUP 04] describes. When looking at the topics of the meetings of the CNRS's staff and information officers between 1992 and 2004, the same queries, questions and requirements arise just slightly later than in the UK. Technical tools, conceptual and organizational models, developments on the TSI market and legal questions were also covered during these meetings. However, when analyzing plans regarding training of the INIST's library staff, two differences instantly spring to mind. These are an extremely high demand in further training in the use of computers, and the use of the English language.

The emergence of personal computing and programs used for administration leads to the need to understand and use tools that are required everyday in the librarian's work environment, such as word processing, spreadsheet applications and databases. Furthermore, it also has an important impact on the relationship between an information officer and the project manager who uses the given information.

English has become the lingua franca of the digital library. All manner of tasks such as analyzing the offers, contacting editors, negotiating licenses, managing online sources, carrying out research, opening download and print documents require different levels of English, especially when dealing with STM publications. Without any knowledge of the English language it is very hard, even impossible, to carry out all of these tasks.

2.8. "He who takes the longest strides…"

Let us now go back to the question asked at the beginning of this chapter: Is there a future for librarians? Google is currently starting to become the world's library. Is there still a role for the traditional library?

Developments in technology linked to the globalization of the market economy of TSI determine the virtualization process of documents.

We have analyzed the development of the librarian's job profiles in the context of TSI by looking at directories, task carried out in the field and the level of training.

It remains difficult to draw a conclusion from the analysis. However, is there only one conclusion? When reading studies on this subject we are torn between fear and feelings of optimism. Optimistic visions include broader skills and increasing responsibilities of the library's staff [LUP 04]. Others see a bright future for librarians. They will become information managers, webmasters, knowledge managers or researchers focusing on new developments within this field [BER 03]. They will manage and administrate systems [LUP 04] or even "turn into cartographers and pilots in the archipelago of knowledge" [BAL 03]. When looking at the down sides of this new trend, fears that the librarian will turn into a legal expert, computer scientist, accountant or negotiator are apparent. In other words, TSI-based professions will increasingly detach themselves from the actual content of information. Are we moving towards new horizons by creating hybrid professions in this field or are we about to commit professional suicide?

We know about the challenges awaiting us but do we also know how to react, how to control this subject which is ours? Do we know how to make things happen [HET 04]?

To escape marginalization, Arnold [ARN 04] suggests a political strategy which involves lobbying by professional associations, a proactive approach towards the scientific community and a strong involvement of committees and interprofessional work groups.

On the other hand, "he who takes the longest strides does not walk" [Tao-te-Ching, Book I, 24]. Blind activism will not be successful. In this case, acquiring knowledge and techniques of other professions will not be an advantage.

Let us start by establishing the features of a digital library. It is mainly a new form of technology that is becoming an increasingly frequent phenomenon [FRA 90] and is based on the subdivision of tasks and the argument that digital libraries are highly efficient and increase productivity (return of investment). It is a model of production without any external control.

To keep or regain the control over our work this new technology needs to be "tamed" and mastered in order for it to be used as a professional tool. Currently, the core tasks of a librarian i.e. cataloging, creating collections, referencing and mediation in research, are increasingly taking place outside the library [SIE 04]. The profession itself is taking a more technologically oriented outlook and moving away from traditional values such as close links with the community and public service. The development of digital libraries ignores problems it might cause for individuals and society as a whole.

Jean-Michel Salaün gives a realistic overview when he speaks of an uncertain future for librarians. However, he also states that librarians still have "very interesting cards" in their hand when playing this game [SAL 00]. Let us play those cards of know-how and traditional values and gain a critical distance from the technological revolution of the digital library that is taking over.

Let us finish this chapter with one last quote from Tao-te-Ching. "With enough practice you could walk without leaving any footprints" [Book I, 27]. We cannot know our future ahead of time; there is no such thing as a path already laid out for us. However, we can ask questions about how useful new technologies are, what impact they have on humans, and what the real interest in using them is.

Maybe we need time to let our cultural schemata adapt themselves to the current overbearing influence of technology. Virtual applications cannot be without any sense or content. "Thirty spokes share the wheel's hub. It is the center hole that makes it useful". Information officers need to fill this center hole in the digital library in order for it to make sense.

2.9. Bibliography

[ABF 04] ASSOCIATION DES BIBLIOTHÉCAIRES FRANÇAIS (ABF), *Débat virtuel: Bibliothécair: quel métier, quel territoire?* http://www.debatvirtuel.bpi.fr/ABF, 2004.

[ARN 03] ARNOLD S.E., "Places and spaces: mapping the library of tomorrow", *The Electronic Library*, vol. 21, no. 3, pp. 6–7, 2003.

[ARN 04] ARNOLD S.E., "Information boundaries and libraries", *The Electronic Library*, vol. 22, no. 2, pp. 110–111, 2004.

[BAL 03] BALTZ C., "Quand la documentation se réveillera…", *Documentaliste*, vol. 40, no. 2, pp. 148–153, 2003.

[BEN 04] BENHARRAT A., "International Summer School on the Digital Library", *BBF*, vol. 49, no. 2, pp. 101–102, 2004.

[BER 03] BERNAT L., "Les documentalistes ont l'avenir devant eux, mais… ils l'auront dans le dos chaque fois qu'ils feront demi-tour", *Documentaliste*, vol. 40, no. 2, pp. 142–146, 2003.

[BOU 05] BOUKACEM C., SCHÖPFEL J., "Statistiques d'utilisation des ressources électroniques: le projet COUNTER", *BBF*, no. 4, p. 62-66, 2005.

[CAV 04] CAVALIER F., "Réseau, publics et territoires: les BU en mutation", *ABF Débat virtuel*, 13 June 2004.

[CNR 02] CNRS, *Répertoire des emplois-types des ingénieurs et personnels techniques de la recherche*, http://www.cnrs.fr, 2002.

[DEG 79] DEGEZ D., "Que va devenir le métier de documentaliste ?", *Le Monde*, 10 March 1979 (reprinted in: *Documentaliste*, vol. 40, no. 2, p. 102, 2003).

[ECI 04] EUROPEAN COUNCIL OF INFORMATION ASSOCIATIONS (ECIA), *Euroréférentiel I&D. vol. 1: compétences et aptitudes des professionnels européens de l'information-documentation*, 2nd edition, Paris, ADBS Editions, 2004.

[FRA 90] FRANKLIN U., *The Real World of Technology*, Montreal, CBC Enterprises, 1990.

[FRO 04] FROCHOT D., "Préface", *Guide pratique Archimag: le droit de l'information*, Paris, Serda, 2004.

[GIC 03] GICQUEL F., "Formation continue: un engagement fort de l'ADBS", *Documentaliste*, vol. 40, no. 2, pp. 104–109, 2003.

[GRE 02] GREENSTEIN D., THORIN S.E., *The Digital Library: a Biography*, Washington DC, Digital Library Federation & Council on Library and Information Resources 2002, 2nd edition, http://www.clir.org.

[HET 02] HETERICK B., "Faculty attitudes toward electronic resources", *EDUCAUSE Review*, July–August, pp. 10–11, 2002.

[HET 04] HETERICK B., "The future ain't what it used to be", *UKSG Annual Conference 2004*, www.uksg.org.

[HIC 04] HICKEL G., "Un référentiel pour la filière bibliothèque", *ABF Débat virtuel*, 13 June 2004.

[INI 99] INIST, INIST 2000 – rapport de synthèse, Internal report, 1999.

[LAO forthcoming] Lao-Tseu, *Tao-te-King* (trad. O. Nyssen). http://www.geocities.com, forthcoming.

[LUP 04] LUPOVICI C., "L'évolution du role des bibliothécaires et documentalistes dans le cadre de la bibliothèque numérique", *Colloque "Bibliothèques numériques"*, 8 April 2004, http://h2ptm.univ-paris8.fr/bn/prez/clupovici_fichiers/.

[SAL 00] SALAÜN J.M., "Du partage des resources à la fourniture des documents: nouvelles relations entre éditeurs et bibliothécaires", in LE MOAL J.C., HIDOINE B. (eds.), *Bibliothèques numériques. Cours INRIA, 9–13 October 2000, La Bresse*, Paris, ADBS Editions, pp. 99–118, 2000.

[SCH 03] SCHÖPFEL J., "INIST-CNRS in France: a model of efficiency", *Interlending & Document Supply*, vol. 31, no. 2, pp. 94-103, 2003.

[SCH 04] SCHÖPFEL J., "The workshops of the annual UKSG conferences 1990–2004: the emergence of the digital library", *Serials*, 17, 3, 243-251, 2005.

[SIE 04] SIERPE E., "Managing the information revolution: library management, the control of prescriptive technologies, and the future of librarianship", *Library Management*, vol. 25, no. 4/5, pp. 177–182, 2004.

This page intentionally left blank

Chapter 3

The Reader Faced with a Digital Library: the Experience of the Pasteur Institute

3.1. Introduction

In 2000 the media library of the Pasteur Institute was faced with an increasing number of digital resources, which are often underrun by users, even though the cost of these digital resources is steadily increasing. Re-designing access to scientific information in order to better adapt it to the needs of researchers and at the same time securing a return on the investments made on those resources is becoming an increasingly urgent requirement.

An analysis of statistics from a study that was carried out among the Institute's researchers showed that the library's web page, which was created in 1998, is used more by "internauts" than by the researchers themselves. The latter mainly use the links to periodicals and bibliographic databases, i.e. the catalog of the multimedia library. When it comes to databases, tools like Ovid allow access to those resources whose access must to be paid for. When compared with the NLM's (PubMed) open access site these resources are not used frequently enough. The prerequisite of learning several programming languages used during a search process as well as knowledge in the different subject areas of every database does not make these database tools very attractive to the researcher. Furthermore, the "lack" of resources that the researchers indicated during the study is a problem as these resources are presented as hits in 90% of all cases, but then cannot actually be accessed by the researcher.

Chapter written by Emmanuelle JANNÈS-OBER.

A tool that allows for centralized access to different sources and is easy to handle at the same time is therefore urgently required. Centralizing services and interfaces with the help of content management systems (ordering books or articles that are not published online, in scientific journals, etc.) also represents added value for the library.

In 2001 the answer to these requirements was discovered, a documentation portal that has been progressively introduced. The final version was inaugurated in September 2003. Statistics on the usage of this portal can therefore only be used to analyze the main trends.

This chapter will give an overview of the documentation portal Biolib that was introduced at the Pasteur Institute. After dealing with content, available services and technical solutions, we will examine how users carry out research on the portal in order to establish if the initial aim of this project was met. In particular, the question of whether a tool like Biolib, which comprises different sources, improves the use of all related resources in the process of research at the Institute is very interesting.

During this analysis we will establish the positive aspects of this project as well as its disadvantages. Technical solutions will be analyzed as well as measures taken in the field of documentation that have been implemented in order to guarantee that all sources are relevant and up to date for the users of Biolib.

3.2. Which services should be aimed at what kind of audience?

In 2000 a new policy on research and science was implemented at the Pasteur Institute. It focused on the co-operation between researchers who worked on related topics. The multimedia library had to adapt itself to this new context.

The target audience is principally internal staff. However, it is also a good idea to establish which groups of the population could be in need of resources and documentation services. All scientific research that is carried out at the Pasteur Institute is related to infectious diseases. The audience is therefore extremely specialized, even if their fields of specialization may vary (e.g. microbiology, immunology, virology, genetics, biochemistry). The majority of researchers at the Pasteur Institute work at its headquarters in Paris. However, the needs of researchers who are localized in branches that are part of the Institute's international network (i.e. around 20 centers can be found in five different continents) also need to be taken into account. Furthermore, the center in Paris is currently moving part of its activities to the suburbs, i.e. to Fresnes.

Therefore, the aim is to create a research tool that enables the user to access several sources at once and is also easy to use. There will be only one interface similar to an online search engine that will enable the user to search many different types of sources at the same time. Access to services could vary according to legal restrictions and contracts agreed between those who produce information and organizations that function as distribution channels.

3.2.1. *Content*

The aim is to access different kinds of resources without changing the search engine that is used.

At the Pasteur Institute the multimedia library comprises internal data produced within the Institute as well as external data to which access is either free or where a payment must be made. The content also needs to be subdivided into structured data (e.g. databases and XML) or data that is unstructured, such as files produced with the help of the Microsoft Office suite, HTML, etc.

Some data is shared between all of the researchers working at the Institute, while other information is reserved to specific groups of users.

3.2.1.1. *Internal sources*

Internal sources are made up of all of the web pages of the intranet and website produced by the Institute. These pages contain scientific information. The content is furthermore subdivided into the following local databases: Bionetbook (a directory of websites specialized in this field), the catalog of monographs and periodicals available in multimedia library, collections of scientific photos, conferences, etc.

3.2.1.2. *External sources*

External sources consist of free websites such as PubMed and Scirus, as well as sites which charge a fee such as ISI web of knowledge, electronic journals and institution websites. Access to Google is also available.

3.2.1.3. *Limiting the access of resources to a specific group*

Any resources that might be of interest to one researcher could be useful to the entire community. New sources which are integrated into the system are generally available to all researchers. When it comes to access that must be paid for, the majority of all subscriptions are financed and organized by the multimedia library which makes this information available to all researchers working at the Institute and, if possible, all other people that work within the Institute's international network. However, it is still possible that a team of researchers may decide to

subscribe to a scientific journal or a database. In that case only the particular team of researchers can access the data, as legal regulations prohibit general access.

Here, the aim should also be to integrate this particular data, which can only be accessed by a particular group of researchers, into the general search engine in order to have only one interface used to access information. This is very important because it standardizes access to information but also offers services that are linked to different sources.

In all cases, information officers integrate, choose and validate the information in order for it to be accessed directly through the portal. The sources suggested by the researchers are analyzed in order to establish whether to allow access to the entire site. As there is the risk of introducing data into the system which is not relevant, the content is analyzed and, in some cases, only parts of the site that are of interest to the scientific community at the Pasteur Institute can be accessed via the search engine. Offering content that is always up to date is another challenge to be mastered by the administrators of this portal.

3.2.2. *Services*

The portal Biolib offers resources and services. The user's identity is verified when accessing the portal. The user is identified and so are his/her prerogatives. This identification process enabled Biolib to provide a personalized service.

3.2.2.1. *Simplified multisource search – different kinds of sources are merged together*

The user selects the required sources and then carries out a search which is merged into one single search engine that automatically goes through all selected sources. The language used is Boolean.

The results are subdivided into different groups according to their source. Furthermore, within these groups the results are listed according to their relevance. The user might choose for the results to be sorted according to a specific date, e.g. all articles published after 20 November.

3.2.2.2. *Multifield search in a single database*

The researcher knows the sources and would like to search one database only; in order to do so several fields enable the user to type in search entries and retrieve information from the database. These fields state, for example, the name of the author, the institution and the name of the publication. This service exists for PubMed, ISI web of knowledge and Scirus. A multifield search interface for PubMed/ISI web of knowledge is currently being tested and will be integrated into

Biolib if the results of different databases can be separated to a satisfactory level. The programming language used in the search process remains the same, i.e. it does not change from that used in the multifield search.

3.2.2.3. *Keeping up to date with new information, information strands and saving a personal profile of electronic search entries*

The Biolib portal enables researchers to save their profile, i.e. sources that they have previously consulted. They can also choose to be informed via email if new documents are uploaded that meet the research criteria defined in their profile.

Furthermore, it is possible to automate search entries for one of several sources for all users or from a specific group of users. In this case the latest data is available from Biolib's homepage through an intermediary service known as "information strands". This service, tested in 2002 and 2003, is now integrated for users who work as a team but are, in fact, in different locations in order for them to be able to access specific sources.

Moreover, all search entries are stored for a period of three months. This enables the researcher to carry out a search that is part of his/her profile history or access the page that corresponds to this specific search entry. This service is of particular interest for databases that are used frequently and whose processing capacity slows down due to the number of users. However, if resources which have to be paid for by the individual researcher or his/her team are entered as a search entry, this information is not stored in the user's profile history.

3.2.2.4. *Access to unabridged text – ordering and reserving documents*

Researchers wish to directly access the information they are looking for. This information consists of, for example, scientific articles. Researchers greatly appreciate the possibility of accessing those articles directly since they are published in a PDF format. However, direct access is not always possible as either an electronic version of the document does not exist or the multimedia library or the respective unit of the laboratory has not subscribed to this particular journal. In that case, the researcher can order a copy of the article that is unavailable for download as a PDF file. If the article is printed, and can therefore be consulted in the library, the order is rejected. If, however, this is not the case, an order will be placed with one of the external suppliers (e.g. INIST, BLDSC, ICIST).

Books that are available in the multimedia library can be found in the online catalog and can be reserved online.

At all times the user has an overview of orders and reservations made on specific documents and the stage of processing these orders are at.

3.2.2.5. *Exporting data into EndNote*

References that are generated during the research process using bibliographic databases are exported into a bibliographic tool such as EndNote that caters for the needs of researchers when publishing articles, etc.

3.2.2.6. *Accessing specific sources*

The integration of specific requirements of a single group is subject to prior analysis by the administrators of the portal. These administrators establish how profitable it is to respond to a certain request, e.g. subscribing to a scientific journal. They check the level of technical difficulty, the interest in exploiting this new source, as well as integrating it into the centralized search engine so that it can be accessed along with all other sources that are already available.

3.2.3. *Programs*

The programs used vary according to the type of source used or the service that is offered. Generally, all programs used are subdivided into two main groups: they are either part of the portal or they enable the user to search databases.

3.2.3.1. *Managing the portal – source control*

Managing the portal, i.e. source control, generally consists of verifying sources and making them available (front office). Enabling access requires the unification of programming languages used during the search process, the presentation of all sources on a single page, and a standardization of the way results are presented regardless of the source.

Source control also manages profile histories, the ordering process of documents, and information strands (back office).

The program used to carry out all of these tasks is Qwam E-content Server produced by a company called Qwam System.

3.2.3.2. *Searching local data and external crawl data*

Local data or static web pages are classified locally and can be accessed using the Verity K2 search tool. A search entry looking for data that is classified within K2 is therefore translated into the programming language K2 with the help of a linking tool that has been developed by Qwam System.

3.2.3.3. *Researching data from dynamic websites*

Searching sites such as PubMed or ISI web of knowledge relies on linking tools developed by Qwam System which enable the user to carry out the search directly on the website without having to worry about different programming languages used on different sites. Integrating a new source therefore requires the development of a new linking tool. Sciencedirect, Embase and many other suppliers of electronic journals are therefore being analyzed and a possible integration of these sources is planned for 2005.

3.2.3.4. *Integration processes in the multimedia library and source control*

The back office is not entirely managed by source control. The orders of copies of articles as well as the order history are based on a program called Ex-Libris that was written by Cadic. This program is also used to feed data into the databases of the multimedia library (catalog and collection of photos). A linking tool between these two programs provides the user with a complete service that can be accessed via one single interface, which remains a rather simple application.

After taking a look at technical aspects we will now see if this tool responds to the needs of the researchers at the Pasteur Institute.

3.3. How are services used?

Statistics on the usage of sources and services of Biolib have only been available as a complete set from September 2003 onwards. The provisional analysis of the statistics needs to be completed through further study on the behavior of information officers who started working at the Pasteur Institute long before the installation of the portal.

3.3.1. *Empirical knowledge and how users carry out their research*

3.3.1.1. *A lack of knowledge in traditional means of documentation*

The choice of simple, to some extent even simplistic, access to multiple sources is based on the following observation: researchers that received formal training in the use of research tools during their academic training represent a minority at the Pasteur Institute. This is very surprising as their job profile implies that these researchers produce a large amount of scientific information. Once they start working at the Institute an even smaller number of these researchers undergoes training on the research tools used in the multimedia library. The portal therefore focuses on users without any formal training in research who generally rely on one single source, such as PubMed. These users can now search several different sources

at the same time by using a universal interface. Have those scientists who did not use traditional research tools in the past been converted and are now relying on Biolib for their research? It is still a bit early to make a final statement on this question. However, the number of Biolib users is greater than the number of people who received training at the multimedia library.

Furthermore, those who are trained in the use of research tools can still directly access servers that are located elsewhere. Using Biolib as a gateway also enables them to obtain details from primary sources since the monobase search engine allows for a very refined search. Is it possible to establish whether experienced users are convinced by the simplified applications of Biolib? When looking at the statistics based on the information accessed on the Qwerty Content Server it is hard to establish whether a search has been carried out by experts or fairly inexperienced users. However, when looking at the user profiles we can establish that a large number of experienced users at least tried Biolib out of pure curiosity. Some of those users have their own profile and are therefore regular users. However, it remains impossible to push this analysis any further.

3.3.1.2. *Users like a simplified search and a high number of hits!*

The fact that the search entries of Biolib users resemble those of traditional Internet users is not surprising. The general search entry consists of one, or a maximum of two terms with one space between them. The use of nouns is frequent and search entries are mainly written in English even though the database is available both in English and in French.

Traditionally, search entries are rather vague at the beginning and are refined when searching for a second time if the number of hits is too high. When using Biolib, Ovide or other sites such as ISI web of knowledge or Scirus, the way search entries are expressed remains the same. A large number of hits does not put off the general user. However, users normally do not like finding the same hit twice in the list of results, even if it stems from a different source. They therefore often ask for a tool that splits up the different strands of information in order to avoid this problem.

3.3.1.3. *Communication increases the use of the portal!*

Biolib was based on the idea that the user does not need any training to carry out research. A bilingual online help desk was quickly installed. However, the use of Biolib is increasing after each presentation organized by the multimedia library. It does not matter how intuitive the tool itself is; the more communication that takes place, the more this tool will be used.

3.3.2. *Some statistics*

The usage of Biolib increased slowly as it took some time to install all of the programs. The first versions of Biolib were prototypes, their services were incomplete and the multimedia library therefore did not want to advertise them. In 2003 the first version of Biolib was officially inaugurated and a campaign was launched to win over disappointed users.

After an improved version had been launched online the number of users has increased tenfold over the past six months. Amongst the regular users were some who received training by one of the library's staff members. However, an increasing number of researchers are working in areas that have not yet been targeted by the advertising campaign. The fact that they are also using the system is probably due to word of mouth.

3.3.2.1. *Who uses the portal?*

Around 70% of all research units or laboratories have used Biolib at least once. Less than 20% have never used it and 25% are regular users. However, Biolib is only used by half the researchers that this service is aimed at, since it is always the same people using the service.

3.3.2.2. *Most frequently used services*

A multisource search is the most frequently used application, probably because it saves the researcher's profile. All sources that could possibly be of interest to the user are saved in his/her profile.

A multifield search of a monobase comes next, but is used much less than a multisource search, as it is used for ordering documents. Other regular users rely on Biolib after searching PubMed or ISI web of knowledge and being unable to download the full text from these websites. Biolib is therefore not used for research but to find references which are then automatically transferred to the monobase's order form.

Before June 2004, ordering documents was subject to a bibliographic search that needed to be carried out beforehand. This is now no longer the case and a strong increase in the use of Biolib can be observed since an order form which does not require this first bibliographic search was made available online. It would be very useful to analyze statistics in order to determine to what extent some users rely on Biolib's ordering service without ever carrying out any multisource or monobase searches.

The number of profiles is impressive, but it would be useful to establish how many of those profiles are actually used on a regular basis. From the very beginning a large number of registered profiles of researchers that never actually use the system could be observed. These profiles were created by users who were experimenting and exploring the system. It is still too early to judge how large this group of profiles is and what might become of it.

3.3.2.3. *The most frequently used sources*

Databases that are part of the multimedia library as well as PubMed are frequently used. Structured information is therefore the most frequently used source. Amongst the websites used, a collection of the latest publications in electronic journals are the most consulted. Internal web pages are the least consulted. This phenomenon was, however, anticipated during Biolib's development process.

Even if it is still too early to scientifically analyze the usage of Biolib, it is still important to underline that Biolib disposes of very efficient statistical tools. These were conceived to enable the multimedia library to adapt sources and services to the needs of users. All data traffic and transactions that are carried out on the portal are stored in a database.

3.4. Current problems

3.4.1. *How to organize the extremely high number of hits*

Users of the portal enter very general keywords. Professional librarians provide the system with very precise search entries that are linked to these very wide search entries. It is therefore a priority to help the user to identify the groups of the most relevant answers. After carrying out a first rather vague search, users should be able to refine their search through simple navigation.

3.4.1.1. *Selecting sources or collections*

Presenting results based simply on their relevance is not a possible solution as the library's corpus is too large for this type of selection. Results were subdivided into groups according to their sources. A multisearch, for example, comprises PubMed, the catalog of the multimedia library, the Pasteur Institute's database of photos, and abstracts of articles in electronic journals to which the library subscribes. All results are separated according to their sources. The user can therefore consult those results by accessing the subcategories, i.e. the different types of sources named above. It therefore enables the user to choose a result according to the desired support (article, reference of a book, photo, etc.). This type of layout is

quite satisfactory; however, this may not remain the case if the number of sources increases.

3.4.1.2. *Selecting data based on categories or clusters?*

Separating results according to their topic is another option. The idea of introducing a mediator when it comes to selecting information was quickly abandoned. The experience with the training of the MeSH thesaurus of the National Library of Medicine has convinced the multimedia library to make a similar investment. However, a classification taxonomy of Biolib's search results happens before the user actually gets to see them. After a long study of tools suggested by programs such as Verity K2 and linguistic corpuses which cover similar fields to those researched at the Pasteur Institute, the multimedia library abandoned this project. Classification of the sources relies heavily on human resources. Employing people, of course, represents an expense that must be covered. This amount of funding has to be justified. Furthermore, researchers continually use new terminology in their search entries. These terms and concepts are often not found in taxonomies and, even if they are, they are not necessarily linked to the same concept.

Another interesting alternative might therefore be separating the results with the help of tools that are linked to clusterization. However, tests on programs that have appeared on the market have convinced neither the information officers nor the IT consultants working on this project. Those programs that were tested in 2003 were difficult to integrate into a complex technical architecture. Some of them were entirely incompatible with the existing system. Others were approved by the information officers who believe that they would add value to the system. The disadvantage of these programs and the reason why, in the end, they were not installed is the fact that their navigation within the search results was rather complex. When carrying out the same search twice, different results appeared. Another disadvantage consisted of the fact that it was impossible to click on an icon that would take the user back to the previous page. Last, but not least, the logic followed by this program was not very transparent and was not always easy to follow.

The multimedia library is currently carrying out tests on technically advanced versions of clusterization programs. The market for this type of programs has stabilized and the programs that are offered can be adapted much more easily than their predecessors. Their integration into Biolib is therefore much more straightforward. However, we already know that clusterization will not be entirely satisfactory. A mixed solution of a semi-automatic analysis of the user's search based on the calculation of clusters at the same moment as the hits are displayed on screen may be the ideal solution. The question of whether there are enough human and financial resources to opt for this mixed solution, however, still remains

unanswered. If the answer is no, the option of using a clusterization tool on its own still remains. This option, however, would put an end to all attempts to introduce a fine taxonomy into Biolib.

3.4.2. *Can the costs be controlled?*

Introducing a portal such as Biolib brings about the challenge of overcoming financial difficulties. In order to control the financial situation, the budget needs to be analyzed in advance. The integration of new resources that often need to be paid for still creates difficulties for libraries and documentation centers. These problems become increasingly difficult to deal with. Open access-based projects will certainly force the producers and distributors of scientific information to review their prices. In the short term, however, it is still very difficult to predict the costs for access to these sites from one year to another.

Furthermore, the creation of a common interface that enables the user to search all those sources at once might oblige the library to buy new programs. Accessing a new dynamic web server might, for example, force the library to buy a linking tool. It might be necessary to extend licenses. As libraries do not have an impact on the development of what is offered on the market, it is not always possible to plan what developments will be required in the course of following year. Buying new programs is not only a strain on the budget when the program is bought, but licenses need to be renewed and therefore represent a constant strain on the budget. Moreover, maintenance of these programs is another burden on the budget. There is no guarantee that a chosen solution will be the right one. Technical developments, economic changes, notably in the policies of publishers, and also financial restrictions have an impact on possible solutions.

In the medium term, the budget of the multimedia library will no longer be sufficient to cover maintenance and the costs for the development of Biolib. As is already the case for subscriptions to electronic journals, all other costs incurred by the multimedia library will have to be shared with its users. This trend is inevitable.

3.4.3. *How to create a new dialog with the user*

The scientific multimedia library initiated the creation of Biolib. However, it remained difficult to make all members of staff, especially those working at reception, engage with this project. The idea that the user becomes entirely independent within his/her research meant some librarians came to question their profession and ethos.

Since the introduction of Biolib and the continuous increase in its usage, it has become much easier to engage with members of staff who were not part of the initial project. The creation of teams that ensure the promotion of the portal created a new form of dialog with the user. Information officers now go to the researchers' laboratories and know about the researchers' work environment and their specific needs. Qualities such as being able to listen and to give understandable explanations are highly appreciated by the user.

To finalize the portal, an online helpdesk has been installed on the library's server. Questions are answered according to the specializations of the staff. This enables the user to benefit from the individual qualities and skills of each member of staff. This service was installed in 2004 and its usage has been continually increasing ever since. The questions asked are very diverse and all answers given are of the same quality. A traditional helpdesk within the library could not have provided this level of service as the quality of the answer would strongly depend on the qualifications of the person working at the helpdesk at the time the user asked his/her question.

In conclusion, the introduction of Biolib has given an overview of the user's behavior regarding digital information. This overview is not at all linked to the idea of whether the user is specialized in a particular field or not. The fact that all resources are found in a single place underlines the virtues as well as the shortfalls of digital libraries. The difficulties professional librarians were faced with became obvious. They needed to change their behavior to adapt themselves to the needs of users. They also needed to familiarize themselves with new software and work hand in hand with IT consultants. Last, but not least, it still remains difficult to keep an overview of the budget as the cost of acquisition is constantly changing. The creation of digital libraries and documentary portals shows to what extent the digital revolution has provoked changes in the behavior of users. These changes are irreversible and have led to the creation of new services within libraries.

3.4.4. *Appendix: Biolib's search interface*

Figure 3.1. *Multisource search*

Chapter 4

University Students' Information Strategies: From Institutional Expectations to Real Uses

4.1. Introduction

21st century French university students are "digital natives". These young adults, born after the invention of micro-processors, are indeed techno-literate: PCs are part of their everyday life[1] and have become essential working tools that they handle with much more ease than their elders. A few years ago it was predicted that young people would make use of electronic resources in ways that differed considerably from their parents' use. The time has come to analyze the true nature of this use.

This study offers a qualitative analysis of students' practices in the domain of online information retrieval. It is complementary to prior studies conducted with students from Lille 3 University [FIC 01; DES 03], and is part of a wider ongoing research project[2]. Our study is based upon the results of various recent international surveys [TEN 03] as well as readership analysis from the Lille University main library. Our aim was to gain a better understanding of students' informational behavior. We deliberately chose not to focus on a particular type of document (such

---

1 A French Government survey shows that 97.7% of people aged 11 to 18 regularly use a computer and that Internet penetration has reached 90% among students. (Baromètre des usages de l'Internet: http://delegation.internet.gouv.fr/).

2 This research project entitled "Culture Informationnelle et Curriculum Documentaire" seeks to identify the knowledge and skills linked to information literacy, from kindergarten to university.

as those that a librarian or information specialist would call "electronic resources") so as to consider electronic resources in the widest sense possible, as whatever a student might use in the course of his work as a digital documentary aid. Furthermore, preliminary observations have shown that it would be inappropriate to distinguish between information retrieval (IR) and other related tasks such as word processing or emailing. Such distinctions would be artificial, since for users the computer is the focal point of various activities and numerous kinds of texts and images, which make sense as a whole.

4.2. Methodological issues

The theme of this book gave us the opportunity to examine the various theoretical and methodological issues surrounding use and user analysis. The main problem was the choice of a relevant focal point from which to conduct our observations and thus to gain insight into computer-based information-seeking behavior. Computers are complex technical devices, but above all, they create new specific work environments that research studies must take into account. The very fact that scholars have found it necessary to invent new words such as *on-screen writing* [SOU 96], *digitized media* [JEA 01], or *scripto-visual devices* [MEU 04] and that they have proposed the term *wreading* [WEI 01] to describe a new hybrid practice, shows that, although these new situations are in some ways similar to older forms (texts, medias) and older practices (reading, writing), they strongly differ from them in many ways. Indeed, the activities that we wish to consider are performed by means of *textual-tools*. They are a complex combination of screen-reading, writing and task execution by means of a technical system.

The characteristic feature of these devices is that they require a textual metaphor to be deciphered in order to activate the system. The texts and graphics which appear on screen might represent tools, as well as processes, data or tasks in progress [DES 04]. Moreover, multiple "strata" pile up on the same screen [COT 04], coming from various sources and belonging to different contexts that must be accurately interpreted. A semiotic-pragmatic approach might thus seem quite relevant to comprehend the way people do things with these distinctive words, given that the resources in question, and the activity which they provoke, can be considered signs in action. However, restricting our study to semiotics would have meant choosing to see things from a predictive point of view, and limiting our analysis to the usability of a given tool, i.e., to the evaluation of a user's ability to conduct a pre-defined task. As users often choose not to do exactly what the original design technicians thought might be done with the devices that they invent [PER 89], we had to envision other complementary approaches. In the same way, our aim was not to limit our observation to the use of technical devices, but to

consider the broader context of use. Thus, it seemed quite relevant to consider the situation, the technical device and the textual metaphors all at the same time.

We thus sought to establish a framework allowing us to consider students as socially situated individuals and to take into account "*socially situated and enacted knowledge, mobilized and constructed for communicational use*"[3]. This approach, which we might call an "ethno-techno-semiotic" approach [SOU 03] does not single out one specific aspect of the situation under study. Similar and simultaneous attention is given to the materiality of the objects manipulated, to the enunciation of both the screen and the situational context, and to related social and cultural practices. Consideration of the technical characteristics of the hardware and software used in a specific situation help to interpret observed phenomena. At the same time, semiological analysis of the screen contributes to our understanding of unexpected behavior.

With this objective in mind, we thus had to determine an appropriate methodology and decide which aspects of the IR situations to take into account. The choice of a group of first and third year students from our university seemed the best strategy, as it gave us access to a considerable quantity of data that would have been otherwise difficult to obtain. This advantage seemed to us to offset the possible "bias" that might derive from our own proximity to the observed population. The question of the observation site also had to be resolved. Given that a student might devote time to IR in many different places, we had to decide whether to conduct a comprehensive study of the information-seeking activities of a few students at home and at university, and elsewhere, at the library and in the classroom; or rather to choose one or two locations closely linked to the tasks we wished to observe.

As this study was to be carried out in the framework of a wider research project planned to last for at least two more years, we decided upon a broad exploratory approach. We thus selected several different observation sites. Furthermore, it seemed essential not to artificially separate IR activity from the other practices to which they were naturally linked, from the students' point of view. We therefore decided to start from the Central University Library, which is traditionally associated with the concept of IR, and then to widen our observations to the other places where IR takes place today. However, we finally obtained much more interesting results by "tracking" a group of students for several days, as we quickly discovered that the material environment in different locations had a great impact upon the organization of students' work.

---

3 "*Des savoirs en actes et en situations, chez des sujets sociaux, dans des communications où ces savoirs sont mobilisés et construits*" [LEM 03].

4.3. Relating use and environment

In fact, the availability of a computer, or more precisely of various types of resources, explains, to a great extent, the way in which students organize their workday at the university. Many of them live a long way from the campus and must thus optimize their time on-site. As the great majority (95%) own a personal computer, they prefer working at home. Yet most of the electronic services, like CDROMs or electronic journals, are available only on the local network, on site at the university. Students must thus determine the location best suited to meet their needs. In addition, certain resources can be reached only from specifically defined areas on the campus. These technical choices clearly show the institutional vision of what students' activities should be within the given perimeter of each access area.

In order to illustrate the way in which students turn these institutional and technical contexts to their own advantage, we chose to study two locations on campus from which students can access electronic resources. Our aim was to discover their "tactics" as Michel de Certeau [CER 90] has described the ordinary uses constructed "against" prescriptive pressures. Our perspective differs somewhat however from de Certeau's, whose model suggests a polarized opposition between the "dominant" and "dominated" and thus tends toward the construction of preconceived scenarios of use, such as, in our case, the notion that students are "subjected" to certain limitations which they manage to "twist" to their own advantage. However, we wished to consider students' behavior as consistent and corresponding to a real "strategy" even if such strategies are actually quite different from what is expected of them by the institution. Use of electronic resources, in the situations that we were given to observe, can indeed be seen as resulting from compromise and from constant adjustment between institutional demands and students' requirements.

The first observation site is a small area, located in the main library, in close proximity to the old "subject" and "author" card catalogs. It is a narrow row of tables with eight computers placed side by side, from which students can access the OPAC. As shown in Figure 4.1, this area was designed to fit the greatest possible number of computers into its small and confined space. PCs are so close to one another that less than the width of a sheet of paper separates two keyboards. The lighting is poor and there are no printers. The close proximity to the old card catalogs is quite interesting for two reasons. Representative of the "traditional" and older method of offering access to the library collections, the catalogs suggest to users that the nearby computers are an extension of their function. In addition, this proximity illustrates the priority given by the librarians to user catalog searches in general, especially since the row of PCs is dedicated to this sole task.

Figure 4.1. *The OPAC querying zone at the University of Lille main library*

For students, who have access to other more convivial work spaces, this is a transit area and none of them stay for very long. Regular users drop by to cross-check data found elsewhere or to note the references of the book they wish to borrow. Others simply "*come to have a look*" as they told us during interviews. Some students do not anticipate the need to write something down, as they do not even have a pen. A free daily newspaper, handed out in the subway, or a library form, is used to jot down a few words.

Computers are polyvalent tools and their presence inside the library can be interpreted in many different ways. In terms of use, we can roughly distinguish between two main types of behavior: in the first case, the person gives priority to the location over the potentialities of the technical object. She links the presence of computers to her representation of the function of the library and her intended use corresponds to the perceived enunciation of the place "saying": "*I am a library*". She gives meaning to the computers as library tools and her project is thus consistent both with this vision and with the students' goals as foreseen by the institution. By contrast, for others, only the computers matter, independently of the surrounding environment or setting. These users consult the computers in order to see "*what I*

can do with them". They bring with them expectations of use which are not necessarily related to the fact that the PCs are in fact "here" in the library. The "perceived" possible uses are thus different according to whether we take into account the setting and, consequently, the institution, as a frame and a prescriptive authority or whether the computer, in and by itself, works as the sign of multiple openings to various types of activities. However, students quickly learn that the "instruction leaflet" pertaining to the computers on the Lille campus is strongly linked to its location. However, the complexity and multiplicity of rules and the heterogenity of the facilities seem to explain the second described type of behavior. The "come and have a look" strategy is prevalent at the university, where students find that institutional rules and expectations are rarely made explicit.

Another location on campus illustrates how students use the equipment at their disposal, adapting their practices to various technical constraints in order to create their own "digital workspace". An open access computer room, equipped with 30 computers, is connected to a network which for security reasons is separate from the main network. The "intra-network" cannot be reached from outside the university. Students are allotted limited storage space that is thus accessible only from inside the campus. Printing is not possible from these computers. These particular choices result from a complex mix of political incentives[4], technical requirements and financial opportunities (French universities are financed through annual subsidies determined by enrolments, as well as, on a more variable basis, through the financing of projects or priorities). Moreover, although a target ratio of one computer per twenty students was set for 2006, actual ratios are still far below this level. As a result, regulations are very strict and students' access to workstations is limited to one hour. These limited sessions make it difficult for students to plan long term work or to immerse themselves in complex demanding projects. However, we found that the students here are quite busy and that they often carry out several tasks at a time: web-searching, emailing, word-processing. A sense of real urgency is manifest in this hub of activity, exacerbated by the presence of impatient students queuing up in front of the door during the busiest moments of the day.

Since students generally have access to computers in quieter places (at home or at their parents' or friends' houses) they do not necessarily make use of this computer room for the PCs themselves. Instead, they come in order to use the particular resources to which they gain access here, either through the local network or by transporting their own files. The computer thus serves as a network relay point: a group project report can be finalized when the contribution of each member has been retrieved from individual mailboxes. A paper is saved on a personal storage space so as to be retrieved and printed in another computer room where a printer is available but where computers have no USB ports. Documents constantly pass from

---

4 Mainly based on the leitmotif "France must fill its technical gap".

one point of the network to another. USB drives and emailing are the main technical solutions adopted for these transfers. The thumb-drives – that are also MP3 players – prove very useful for sending files from one computer to the other in the same room even though the PCs are connected to high speed networks. They thus constitute a kind of "digital backpack". E-mail has also taken on the innovative function of a "document management system". Students open various accounts so as to benefit from larger storage space; they create thematic mailboxes and distribute their messages among them. Sorting and finding functions thus make it easy to locate a particular file according to criteria such as date, subject or sender. Even multi-versioning is made possible by the "send again" option. Students have thus constructed remarkably efficient working spaces and file management systems, whose functionalities are not so different from the ever-anticipated "digital campus" political projects.

These practices present a significant evolution with regard to writing and, to a greater extent, university work. Firstly, students now depend to a great extent upon a technical tool which is not as portable as pen and paper. Secondly, they now no longer perceive documents in a material, tangible way. Writing has become more transient, diffuse and somehow less stable than when it was inscribed on a sheet of paper. The loss of a file is an everyday phenomenon and the mere idea of completeness seems superfluous as documents are always in progress, with diverse versions distributed on various storage spots. Above all, the distinction between what is "mine" and what is "theirs" is sometimes rather blurred. This issue goes far beyond the scope of this paper but we hypothesize that the new materiality of texts changes the ways in which students consider the knowledge they produce, consult, or exploit. When displayed on the same screen and sometimes inside the same word-processor frame, texts become so formally similar that the particularity of one's own contribution seems less apparent. Handwriting is effortless [PIO 05] as compared to typewriting, but our relationship to the text and to knowledge is certainly quite different according to whether you "cut and paste[5]" or copy out in writing the same piece of text.

4.4. Resource legitimacy

In the same way, the computer as an environment constitutes an intermixing of various metaphorical, physical and textual "locations" (the library, the typewriter, the shopping mall, the diary, etc.) which belong to clearly separate social spheres.

---

5 "Cut and paste" practices show an evident lack of respect for copyright issues and intellectual property. Teachers acknowledge that it is an increasingly frequent practice among students, yet the very notion of the illegality of this practice seems unclear to them. A recent survey shows that more than half of "cut and paste" sources are not quoted [Six Degrés – a company who produce software to detect plagiarism on the Internet].

This complexity adds considerable difficulty to the students' IR task. The boundary between what a professor might consider a valid source and what is unacceptable is not always easy to perceive. The university's standards of rhetorical and formal quality for submitted work are high and students must learn how to meet these standards. However, the texts and information found by students on the Web are only seldom the work of recognized authors. Editorial or scientific "filters" do not operate online and technical algorithms or economic interests often prevail in the selection and visibility of texts. At the same time, political, economic and even academic pundits, after having repeatedly announced the advent of the information society, now proclaim the age of knowledge. These enthusiastic and idealistic pronouncements promote the idea of instantaneous, intuitive and direct access to all world knowledge. According to this vision, the need for mediation seems outdated and superfluous.

Students actually do use computers and the Internet in almost every domain of their private life: to download music, to talk to friends, to publish or exchange information, to buy online, etc. They are thus predominantly techno-literate and quite able to "*find things on the Web*". Yet they also have a clear perception of their professors' demands and know that they cannot "*only refer to online sources*" and that "*there should be books*" on their bibliography, as two students explained. In this particular case, they had all the information they needed to finish their report but were still searching the OPAC for books. They are faced with the contradiction between strong intellectual and formal demands and social discourse tending to promote the idea that all the information they need is at their fingertips. Students must cope with this tension between hard to access but legitimate source material and easy to find but much more undifferentiated texts on the Web. Although they think that they have mastered online search techniques, students must in fact combine two types of skills: firstly, skills that will help them select, from among the numerous texts found through search engines, the ones that match their professors' requirements, and secondly, skills allowing them to master the specialized tools and approaches needed to access most scientific resources.

This often proves a hard task for beginners, as they only have an approximate knowledge of the scientific domains when they start off at university [COU 97]. However, professors and instructors implicitly consider that most of the skills and knowledge needed to succeed in an academic environment are already obtained by students upon their arrival [ROM 00; MIL 03]. For example, high-school students generally have only very limited skills in word-processing. However, as they find when they begin coursework at university, their professors require summaries, indexes and other advanced forms of textual layout requiring advanced skills. These skills are thus usually learned from friends on an informal and incomplete "hit and miss" basis. The same observation can be made in the field of IR. University lecturers and professors usually distribute bibliographies at the beginning of their

courses, with sources intended to provide wider or more complete coverage of ideas discussed during the class lecture. These bibliographies are structured according to strict norms, but very few teachers actually take the time to present them. Students are thus supposed to comprehend the subtle formal differences between a monograph[6], a scholarly journal or an edited book. Little or no advice is given to students to help them look up or locate documents. For example, when a lecturer includes a scholarly article in his bibliography he does not tell students not to try and find it by searching the OPAC since individual journal articles themselves are not searchable.

Therefore, students turn to what they know best: Google®. They particularly appreciate the fact that whatever words they choose for their Google searches, their success rate is much higher than with structured library resources [DES 06]. Furthermore, as they navigate from site to site, they manage to collect pieces of information that turn out to give them a better understanding of their subject than the list of references they are expected to use as points of entry into a particular domain. On the Internet, they find pieces of texts, extracts, abstracts, book reviews, comments and opinions, all of which provide them, very quickly and at a low cognitive cost, with enough information to meet their professors' demands[7]. As an extreme example, a student explained that he had "*read*" one of the recommended books on the Internet. The book itself was not available on the Net, but he considered that what he read about it must be more valuable than anything he might have been able to gain from actually reading it himself.

The issue of source credibility often arises next when students have to cite their sources. Though they sometimes wonder how to cite a Wikipedia® entry, most realize that only certain types of documents can be included in their bibliographies. Sources found on the Internet such as the content of a shopping basket filled in a Cyber-library or the advice of a "blogger", need to be "validated". It is at this point that they turn once again to the institution, in order to confront their Internet findings with the credible and authoritative resources of the university OPAC system or library holdings. Two students, who had to write a paper about "*Second Empire Parks & Gardens*", had learned while searching the Web that "*Haussmann had something to do with*" the design of Parisian parks at that time. Their search finally yielded no convincing results because the books they found interesting were located in a research library inaccessible to first year students. Despite this problem, these students firmly believed that Haussmann was an author, because his name appears on the indexed OPAC subject list as an author's name. This example demonstrates

---

6 The term "monograph" while it would seem to be transparent, is never used in high schools.

7 A nationwide survey of students and academics in the Netherlands found that two-thirds of respondents believed that searching the Internet yielded enough or more than enough information [VOO 99].

the difficulties faced by novice students in their attempts to conduct expert searches. The semantic complexity of the screen pages adds to the technical difficulty of the query process. However, more than anything else, they lack the specialized knowledge required to formulate a pertinent query or to interpret the codified and norm-specific information displayed on the screen [DES 00].

For experts, the OPAC is an extremely useful tool, yet because of its rigorous and strict structure[8], it constitutes an obstacle for novice students. These students often do not even make an attempt to understand the OPAC system, since their efforts would not necessarily even "pay off". The two empty-handed students finally turned to the librarian, who guided them to the right shelf after they showed her the sheets of paper printed from the Web.

4.5. The evolution of the figure of the "third party"

In addition to traditional knowledge mediators such as teachers or librarians [BAR 96], other "third party" figures are now beginning to emerge. In the preceding example, a dialogue took place not only between the students and the librarian, but also between the students and the authors of the web pages used to find the books[9]. In this particular case, the websites in question are institutional sites (the City of Paris and the Napoleon Foundation) and the articles were signed. However, participatory projects have become more and more common on the Web. Ordinary citizens have entered into the sphere of public communication and now take on roles which previously belonged to media specialists. Thierry Lancien has emphasized the waning role of journalists [LAN 03]; this phenomenon also characterizes editorial and documentary activities. Projects such as Wikipedia® challenge the very principle of authoring and editing as intellectual and individual endeavors. Apart from these ambitious projects, numerous less visible initiatives allow "ordinary" speech to be published, in areas previously dominated by experts and professionals. Online bookstores now display comments and evaluations posted by anonymous readers and anyone can become a literary critic. Marketing strategies transform online purchase sessions into advice offered to potential buyers interested in the same book as another customer: "*Customers who bought this item also bought…*".

Blogs and discussion groups have also contributed to the emergence of new types of experts. Students who join online tutorial groups develop their own strategies for identifying the most "credible" contributors: they judge the answers by the nickname of the writer and feel honored when one of the tutors of the discussion

8 French librarians use the adjective "*pure*" to describe the bibliographic data entered in the OPAC. Conceived by library professionals, this is the only data which can be added to the French Universities Collective Catalog (SUDOC).

9 These pages were extracted from the "City of Paris" and "Napoleon.org" sites.

list deigns to answer them. The technical or pseudo-technical expert also plays an important role in this complex environment. Those whom we met admit that they mostly learned "*on the job*" and all have a referent, someone who controls the device and introduced them to various "*tips and tricks*".

However, the social and intellectual status of these "experts" is unclear. Over past centuries, lawmakers and editors have created various processes of accreditation and validation in the traditional publishing world: typographers, compositors, printers and bookbinders not only produced books but also took on an important role as knowledge mediators. Concern for readers led publishers and editors to seek innovation and to select and adapt texts for various types of readers. At the other end of the editorial chain, booksellers, newsstand keepers and librarians have also had an important mediating influence upon readers. These editorial and commercial processes contributed to the social "accreditation" of different professional categories: writers, journalists, scholars, etc. whose writings are subject to ethical standards. However, most of the new publishing activity on the Internet occurs from outside these traditional processes and the "right to write" is now potentially granted to anyone. Students are thus confronted with a difficult task as they must evaluate by themselves the relevance of the texts that they "glean" from the Internet and the credibility of their authors. Our research brought to light the "certifying" role played by the institution [DES 06]. It appears however that the people we observed rely on numerous and extremely varied elements[10] to identify the kinds of texts they are confronted with. We are currently planning a new study that aims at identifying these elements as well as the skills and knowledge used to relate various texts to their social and editorial contexts.

4.6. Conclusion

Information retrieval strategies have significantly changed as digitization has impacted every aspect of the publishing and editorial chain. Our observations confirm that the use of information technologies dramatically changes students' relationship with knowledge and information sources. Not only do computers modify their work environment, but they also transform the way students appropriate knowledge and acquire new skills. As new actors now enter into the field of documentary mediation, the scope of IR strategy analysis has to be broadened.

---

10 A French research group has discovered that high-school students used Google-Images® to find potentially interesting sites, and that their main pertinence criterion was the quality of these images.

Furthermore, the technical equipment used by students for their university work is typically the same used for numerous other everyday tasks. Multiple social spheres and activities, that once were totally separate, now converge and mix on computer screens, contributing to the creation of a complex semiotic system. These factors are of great importance in the field of use analysis and should be taken into account in establishing research methods and approaches.

It clearly appears that IR cannot be separated from other activities performed at the same time, since these activities make sense as a whole for users. Similarly, user analysis benefits from a holistic approach in which the observer does not focus on one particular type of electronic resource or on a particular location. Nor should we assume that students themselves make a clear distinction between various types of sources and resources which can appear related, if only because they are accessed using the same equipment and procedures.

4.7. Bibliography

[BAR 96] BARATIN M. JACOB CH. (eds.), *Le pouvoir des bibliothèques : la mémoire des livres en occident*, Paris: Albin Michel, 1996.

[CAL 88] CALLON M. *La science et ses réseaux. Genèse et circulation des faits scientifiques*, Paris: La découverte, 1988.

[COT 04] COTTE D., "Ecrits de réseaux, écrits en strates", *Hermes*, no. 39, 2004.

[COU 97] COULON A., *Le métier d'étudiant. L'entrée dans la vie universitaire,* Paris: PUF, 1997.

[DEC 90] DE CERTEAU M., *L'invention du quotidien, 1. Arts de faire.,* Paris: Gallimard, (Folio), 1990.

[DES 03] DESPRÉS-LONNET M., "Le couple dispositif/pratique dans les échanges interpersonnes", in Souchier, E., Jeanneret, Y. and Le Marec, J. (eds.). *Lire, écrire, récrire : objets, signes et pratiques des médias informatisés,* Paris: BPI/ Pompidou Center, p. 352, (studies and research), 2003.

[DES 04] DESJEUX D., *Les sciences sociales,* Paris: PUF, 2004.

[DES 06] DESPRÉS-LONNET M., "Les étudiants et la documentation électronique", *BBF*, no. 2006/2, 2006.

[FAU 04] FAURIE I., "Les usages d'Internet des étudiants : Facteurs affectant l'intensité, l'orientation et la signification des pratiques", *L'orientation scolaire et professionnelle*, vol. 33, no. 3, 2004.

[FIC 01] FICHEZ E., LE MAREC J. and DE VAGHELAER N., "Usages des réseaux en milieu universitaire : quelques articulations entre savoirs non-formels et normes académiques", *Recherches en communication*, no. 16 (Interfaces sémiotiques et cognition), pp. 47-64, 2001.

[JEA 01] JEANNERET Y., "Informatic Literacy : manifestations, captations et déceptions dans le texte informatisé", *Spirales*, vol. 28, "Nouveaux outils, nouvelles écritures, nouvelles lectures", pp. 11-32, 2001.

[LAN 04] LANCIEN TH., "La médiation journalistique au risque d'Internet", *Médiations et médiateurs, Revue MEI,* no. 18, 2004.

[LEM 03] LE MAREC J. L'usage et ses modèles : quelques réflexions méthodologiques, *Spirale*, no. 28, pp. 105-122, 2001.

[MEU 04] MEUNIER J.-P., PERAYA, D., *Introduction aux théories de la communication : Analyse sémio-pragmatique de la communication médiatique*, Brussels: De Boeck (Second edition), 2004.

[MIL 03] MILLET M., *Les étudiants et le travail universitaire,* Lyon: Presses Universitaires de Lyon, 2003.

[PER 89] PERRIAULT J., *La logique de l'usage : essai sur les machines à communiquer,* Paris: Flammarion, p. 253, 1989.

[PIO 05] PIOLAT A., OLIVE T. and KELLOGG R.T., "Cognitive effort of note taking", *Applied Cognitive Psychology*, no. 19, pp. 291-312, 2005.

[SOU 96] SOUCHIER E., "L'écrit d'écran, pratiques d'écriture et informatique", *Communication & langages*, vol. 107, 1996.

[SOU 03] SOUCHIER E., JEANNERET Y., LE MAREC J. (eds.), *Lire, écrire, récrire : Objets, signes et pratiques des médias informatisé*, Paris: BPI/ Pompidou Center, (studies and research), 2003.

[TEN 03] TENOPIR C., *Use and users of electronic library resources: an overview and analysis of recent research studies*, Report of the Council on Library and Information Resources, Washington D.C., 2003, http://www.clir.org/pubs/reports/pub120/pub120.pdf.

[VOO 99] VOORBIJ H.J., Searching scientific information on the Internet: a Dutch academic user survey, JASIS, vol. 50, no. 7, pp. 598-616, 1999.

[WEI 01] WEISSBERG J.-L., "Figures de la lectacture : le document hypermédia comme acteur", *Communication et langages,* no. 130, pp. 59-69, 2001.

This page intentionally left blank

Chapter 5

The Digital Spirit:
Digital Libraries and Democracy

Digitization represents much more than a simple technique. Due to all possible forms of usage and applications it can be seen as a cultural phenomenon. Other than technical possibilities that have been opened up by digitization, the general public has now also come into contact with some applications that have become increasingly important. We can currently observe certain ways of thinking and behavior that indicate the beginning of an era known as the "the digital spirit". We may ask ourselves what role the library will play in this new world and if this process of digitization will enhance democracy.

5.1. Books and libraries function as an objective spirit

A long time ago historians and anthropologists established the fact that readers "are never confronted with texts that are entirely abstract, completely detached from ideals and free of all reality" [CAV 97]. Texts are always found in a material framework, i.e. the support used to publish them, and these materials reveal a lot about the historical context and the techniques used in the editing and printing process. This shows that the material support used to publish a text is far from being entirely neutral, it has a substantial impact on the meaning of the text. [EIN 91, GOO 96, KEN 91] are the essential texts published on this topic. This fact allows for a detailed distinction of different realities; however, it is questionable whether this impact on the meaning of text is entirely positive as the identity of texts seems

endangered by the material support used to publish them. Changing the support used to publish the text might therefore change the text since the support has an impact on how the text is read and how it can be received. Texts which transcend any form of material support thus do not make sense. An objectified spirit and an objective spirit are confused with each other ([ARO 91] quoted in [DES 96]) and this confusion paves the way for a relativist position.

Let us come back to the nature of books and the library that contains so many of them. A book is, once it is printed, irreversible in its materiality and typography, in other words its physical state is stable. A book is a piece of work that springs from the mind and represents within itself a certain ideal that lasts throughout time and endures different forms of support which function as distribution channels. A written piece of work consists of a set of thoughts that are organized as a discourse and can be referred to as an "intellectual unit". This intellectual unit can be found in the volume of a book. The material identity is essentially subordinate to its spiritual identity. The text becomes part of the objectified spirit once it is integrated into a materialized object and therefore, in principle, is accessible to an undefined number of readers. The text, however, remains part of the objective spirit as it has an impact or signifies something for society as a whole, which of course also includes the individual.

A book never just exists on its own. It is not conceivable on its own and is always part of a plurality of books, i.e. it is linked to a multitude of other works/books. It describes a group of objects or, in other words, texts. It is, furthermore, inseparable from a set of assets such as its use and its application. These assets determine its functions. Books are therefore a social institution. Their existence is linked to a set of social players of which the most important are authors and readers. These players cannot exist without the transmission of the necessary skills, i.e. reading and writing. Without schools as social institutions there would be no such thing as publishers and readers of books. Last but not least, reading and writing needs to be attractive, in other words there must be convincing reasons for the individual to take these skills on board. This chain of complex interdependence represents the social aspect of books.

As libraries guarantee access to books, they represent an important link in this chain. Every library has a material as well as an intellectual aspect to it. Is the library less dependent on its material aspect, i.e. the fact that it contains books, than the book is dependent on its material support? Strictly speaking, a library is simply a place where books are stored and put in a certain order. It is a building created explicitly to store volumes of books. Of course, this definition does not include the essence of a library. Defining a library by stating the number of documents that are stored within it would be just like characterizing the Louvre, a famous museum in Paris, by its number of rooms and paintings. What is the real definition of a library? What makes

the library a library and not simply a place used for the storage of books? Of course, a library is not simply made out of the books that are stored within it. A library is an organized collection of books and other texts. It is based on the principle of storing all of its documents in an organized way based on a structured intellectual reality. The material reality, i.e. the building, rooms, stack-rooms, shelves and tables, is also subject to the intellectual intention of organizing the library. A library is furthermore defined by all of its possible uses and practices that are at least partly determined by the procedures that need to be followed to access documents. The library is a place where documents can be consulted by following an established system.

As these criteria are rather vague, an attempt will be made to define the library without taking its material characteristics into account. A library is essentially an institution that can be defined sociologically. It can be described as "incarnated imaginary signification" (Castoriadis), "*institutions du sens*" (Descombes) or "objective spirit" (Hegel).

5.2. The symbolic value of books stored within a library

A library can first of all be compared to an encyclopedia, maybe a better term for it would be an "encyclobrary". The intellectual organization of a library comprises all the knowledge stored in the library and therefore establishes the way in which the books and documents are stored and organized. The spirit of a library cannot be separated from the intellectual spirit that organizes documents. This form of organization allows for access to precise knowledge by following the way collections are globally structured to eventually find the place where the individual unit of knowledge, i.e. the book or the documents, is stored.

These classifications of collections that stem from intellectual organization are of a pragmatic nature. While responding to scientific, i.e. philosophical systems of classification, they are inseparable from practical aims. A library's spirit is therefore also made up of the techniques required to access documents such as reading, consulting, copying, reproducing, annotating etc.

Apart from this practical dimension, books have reached an almost divine status in society and this sacred character still has an impact on today's society. The identity of a book is not only intellectual but also symbolic. In its material form it represents the aspirations of the human spirit which are unity and universality. As it is a closed unit within itself, it emphasizes a doctrine that reflects a piece of reality, i.e. a piece of the cosmos. It has a beginning and an end and neither can be changed. The path from beginning to end has already been established and makes the book a stable refuge within reality on whose discretion we can always rely. For a long time there was only one book, the Bible. The Bible is also a brilliant example of a book

whose principles are different from all others. When referring to this book the analyses of [JOH 94] will always be the basis of the arguments brought forward within this article. In a rather secularized form, today's society has preserved this cult of books, as books are perceived as a support for great texts that are created by human minds.

This symbolic value is therefore also applied to libraries as a public monument, a temple of a non-denominational cult that worships knowledge and culture as the basis of the existence of honest humans and modern citizens. This is at least the ideal of the republican library which is inspired by the philosophy of enlightenment.

5.3. How can the project of a digital library be realized?

What happens to this intellectual unit of heavy symbolic value if we move from a paper support to a support that is electronic, from a written text to a digital text? Does this technical change imply a modification of the characteristics of texts as well as the way they are read? Does the electronic transmission of texts impose new ways of reading texts as Chartier states [CAV 97, p. 35]? It is undeniable that the new usage of texts is closely linked to the emergence and development of the Internet. On the other hand, it is very unlikely that the invention of digital documents is the sole reason for this new trend. As previously mentioned, the material dimension of libraries is less important than their intellectual aspect. Is it not the users' new expectations of written documents that lead to the increasing success of digital documents?

At first glance, a digital library is a library that is dematerialized and therefore free from all material restrictions linked to paper documents. According to Chartier this type of library is a "library without walls" that "gets rid of the distinction that until now was always present, between the location of the text and the location of the user" [CAV 97, p. 38]. The digital library is a virtual space; its data consists of digital codes that are saved in the memory of a computer and are put on screen as a text the user can then access and read. Documents can therefore now be accessed from a distance. The digital revolution is thus linked to the revolution of telecommunications[1].

If a book loses its character as a modifiable object, how can the intellectual unit of a text be preserved? Is there not a risk that this virtue will be lost? How does the

---

1 According to Jacquesson and Rivier "digital libraries get rid of two restrictions. They cut the link between the content and the support which were strongly linked to one another beforehand. As they are part of the architecture of telecommunications, they change the link between the center of on-going events and areas that could be referred to as suburbia, they make notions such as time and location less important" [JAC 99].

identity of a library manifest itself if there is no longer a building that symbolizes it? If the library is an institution with its very own spirit, this spirit needs to be expressed in one way or another. It needs to lead to the creation of certain techniques applicable to the exploitation of the libraries' capacities or, in other words, take the form of a mediator. If the materiality of a library disappears, how can it be part of a place that is situated in a visible space? How can this type of library obtain a physical character, i.e. where does its material body come from?

With texts becoming less dependent on their material support, their intellectual and symbolic integrity may become weaker and even come under threat (Chartier). Generally speaking, there should not be a problem in creating units of texts that can be assembled in coherent collections of digital documents. The traditional organization of a library, which was based on the organization of books, would be replaced with programs that organize digital documents. Material indicators will be replaced with symbolical indicators. As described in [JEA 00, Chapter 4] and [JEA 01] the digital world generates a semiology with its own visual orientation on screen which is linked to symbols that are part of a specific system. This semiology will replace the sensorial and kinetic orientation in a physical space. Transferring documents from a paper support to a digital support does not normally pose any problems when it comes to the organization of documents. While the computer's memory and its organization are invisible to us, texts still remain accessible through classification and indexation. There is no technical obstacle in this instance. The main principles of organizing knowledge and the culture of printing documents can be transferred to digital documents, i.e. obtaining similar functions within the digital world.

In principle, nothing opposes the transfer of the rules that have been established for the organization of a library to texts that are published using a digital support. In both cases, it is possible to classify and index documents in such a way that they can be accessed by consulting a catalog. The distinctions between descriptive instructions and documents, between primary and secondary information, as well as between textual data and metadata enable the user to access documents systematically. These two different levels of information are introduced in a library. On a cognitive level, there is not that much of a difference between finding the right shelves to locate the required book or clicking on an icon that will take us to a document. Overall, there is not a great difference between these two aspects. The idea of a library is therefore not being lost.

Orientation in a digital library, however, remains rather indirect and abstract and could therefore be classed as a skill in itself. As digital libraries are stored within computer memories and organized in the form of an encyclobrary, finding documents within them requires an understanding of abstract symbols and icons as well as tangible skills of orientation within the library. The feeling of satisfaction the

philosopher Alain obtains when looking at the volumes in a library and reading their titles or the text inside, following the path the author laid out for his/her readers, examining a book's characteristics such as its size, color, its position and the state it is in, finds no equivalent in a digital library. Until now, the usage of texts and the usage of the library's physical space have been far more frequent than the usage of audiovisual media.

5.4. Digital libraries are actually very rare!

The obstacle in the creation of digital libraries is based on the difficulty of "establishing a text" (Chartier) as a digital support. In the long term, technical obstacles that might be encountered when creating a text using an electronic support will be solved.

Other obstacles will, however, still exist. A universal library that can be accessed by everybody is a utopia of the digital revolution. If it can fulfill two conditions this ambitious project can, however, be carried out online. First of all, households would need to be equipped with computers that are connected to a telecommunications network. Secondly, documents that currently only exist in a written or printed form would need to be transferred to an electronic support. This project therefore requires a program that is large enough to digitize all of this data. *Gallica,* the largest French digital library which was inaugurated by the national library of France (BNF) in 1992, currently only comprises 70,000 different objects out of which only 1,250 are written texts.

Digital libraries are still very rare. This is partly due to economic restrictions and partly because digital libraries are not focused on by the media. The digital challenge is nothing more than a simple transfer from printed documents to a technology that imitates them. The question is not how the usage of technical developments could lead to the integration of libraries into new media, but how a cultural project is influenced by the current technological revolution. We therefore need to take a look at *the digital spirit.*

5.5. Technical supports and new ways of reading

New media apparently leads to new ways of reading and writing texts. These new practices taking place before us are part of a cultural revolution. Apart from traditional forms of reading and writing, a new way of thinking is currently developing. This new type of culture that is sometimes expressed in a very militant or even prophetic way can be defined as the digital spirit. Those who advocate this new trend are convinced that digital technology will change the modes of

distribution and access to knowledge and culture in the education of children as well as in the field of further training for adults. Society will undergo a revolution and progressively transform into an "information society".

Which practices can currently already be observed? In what way, compared to the traditional culture of printed documents, do they represent a new and different relationship with knowledge and culture?

Printed texts impose a uniform and unique way of approaching a text: they must be read in a linear fashion. The reader therefore remains rather passive when reading a printed text. Reading a book has always consisted of following the discourse from the beginning to the end by turning every page and reading the entire text from the first page to the last. Is this true or is this image a strong reduction of reading capacities that corresponds to inexperienced and rather clumsy readers? In reality, there are many different forms of reading and all of them find a suitable form of physical representation in the traditional art of printing.

Nothing stops the reader approaching a book in different ways, choosing their own rhythm as well as reading only certain parts of the text. A dynamic reading process of a text slows down during difficult passages and speeds up when the text becomes easier. Experienced readers who no longer have difficulties with decoding the individual letters, anticipate the direction the text will take, i.e. what idea will come next. Just like the philosopher Alain, we may advocate the pluralized form of reading that is adapted to the specific type of text and its language. The reading process is either fast or slow, superficial or concentrated. The reader follows the author's thoughts without being a slave to them [ALA 86]. From the 1960s onwards, methods of an active and also fast reading process were analyzed in the fields of physiological and cognitive processes [RIC 86].

On the other hand, the editing process consists of a number of specific tools that allow for orientation when reading. These tools have constantly been improved throughout history. Tables of content, abstracts, indexes, footnotes, pagination and concordances are just some of the features that help the reader when approaching a text. What Genette refers to as a "paratext" is made up of parts of the text that are separate from the main body. The paratext shows the plurality of a text and indicates its usage. Printed texts therefore are not as rigid as the advocates of digital technology would have society believe. Vandendorpe put the idea of linear reading processes into perspective by emphasizing new aspects of tabulation. "Written texts allow us to escape linearity under the form of a codex that enables us to explore the sophistication of space. A book is enriched by several elements of tabulation" [VAN 99].

If printed written texts are just as well structured on a synchronic level as well as on a diachronic level, what are the new possibilities opened up by the digitization of texts? Every line of characters can potentially be accessed directly. The reader does not need to look for this line of characters. There are multiple ways to enter, access and exploit a text. A large field of linguistics that includes style, semantics or syntactical aspects is now open for analysis. During the era of printed texts, this process would haven taken a very long time. Possibilities of objectivity enable the researcher to move away from the level of intuitions they never strayed from previously. An increase in IT tools will open up new possibilities for analysis that have never been seen before in the fields of literature and other texts.

Furthermore, digitization allows for the modification of texts. Each digital text can not only be accessed and exploited, but can also be quoted, analyzed, cut, reconstructed and a concatenation *ad libitum* with any other text can be carried out. Readers can integrate their own texts into those of other authors and vice versa. Reading and writing processes are becoming increasingly linked to each other and will eventually merge into one process, which means that readers will finally discover that their status has improved and they have *ipso facto* joined the league of authors[2].

The essential novelty of digitization consists of the infinite mastery of texts. The reading process of a printed text is necessarily subdivided into two phases. In the first part of the reading process, notes are taken or observations are made on the text. During the second part, the actual writing process, quotes and observations from the text are integrated into a new text. These quotes and observations underline the intention of the reader who has taken on his/her position as an author. Digitization offers new possibilities that make the composition of a text and the integration of other texts much easier. The two different levels of reading mentioned above are therefore merged into one. Traditional operations undergo further progress and are merged during the text production process. In its most sophisticated form, this process has given birth to a project known as computer-assisted reading. This project caused much controversy but was never actually realized [STI 91,VIR 93, ZYS 99].

However, new forms of reading and writing will not only develop because the technical possibility exists. Other more decisive factors are needed that support education and meet society's expectations in this field. Digital applications and techniques that are transmitted by the usage of the Internet correspond to a

---

2 According to Chartier the reader becomes co-author of an electronic text. "In front of a screen, the reader takes part in a writing process carried out by several players. He/she is at last in the position of creating a new text out of fragments that can be cut and pasted freely" [CAV 97, p. 37].

sophisticated logic that was created by the project managers of computer-assisted reading projects or HyperNietzsche [IOR 00]. Computer-assisted reading and writing require new skills that differ from those required during the reading/writing process of printed documents. These new programs are aimed at texts that become increasingly complex. However, advertising on the Internet targeting the general public states that these will make the consultation and exploitation of texts easier. The values expressed in the advertising campaign are high speed, conviviality and interactivity. This new product makes the knowledge acquired by mankind accessible to everybody. The focus of digitization has shifted towards the Internet. This new technology insists on its democratic character, i.e. it is linked to a second phase of enlightenment. Making knowledge accessible to everybody should enable mankind as a whole to progress. The analysis of a new logic in reading processes and textual practices is now taking place on a political and sociological level.

Except from predictions made by specialists in information and telecommunications and some futurologists, the new possible uses of digital data do not correspond to the applications that can be observed within the closed circles of university researchers. This is why we will move on to other questions. Other practices will be analyzed to gain an insight into the cultural significance of these new trends.

5.6. Two different types of logic within reading processes

The legitimized model of a reading process was traditionally linear. The reader carefully read the entire text. This was due to the authority of great authors and classical works. Their reasoning and arguments were to be read with respect for their work and their thoughts.

It is impossible to only read randomly selected extracts from a book written on mathematics. The same applies to all other texts that rely on a linear form of reasoning and are therefore based on a very solid structure. Great books cannot be cut and pieces of the text cannot be read randomly. An extract of the work *Phénoménologie de l'Esprit* of one of the great commentators on Hegel could be taken as a classic example for a reading process. "…it is of great importance that everybody finds their own rhythm when reading. It is difficult to find this rhythm as it does not only depend on the individual's capacities, but also on a certain number of objective criteria. In fact the 'sense' of the phenomenology is aimed at familiarizing oneself with the sequences of the different figures contained in it (…). Whether the reading process is too fast or too slow, the integration of these particularities cannot be carried out as it would be a hasty solution, or one that remains diluted and therefore cannot be 'carried out'. What follows is a very common and desperate experience. The text that is read only consists of sequences

of words that lack a meaning as a whole. There is no point in stubbornly continuing to read the text. The reader needs to go back around 10 or 20 pages and start the passage again…" [LAB 79].

The cultural heritage of literature also relies on the reader's effort to follow the author's thoughts in all of their twists and turns, as well as to analyze a text in all of its details. This exercise makes the reader sensitive towards Lefort's suggestion on Machiavelli's approach known as "the work on a text" i.e. continually coming back to the author's ideas in order to be able to deal with difficulties of expression and the use of language. The reading process cannot only be focused on the object. Attention also needs to be paid to what becomes the subject as the subject thus thinks and expresses itself [LEF 72].

Today a different logic in the approach to texts has taken over. This logic can be characterized in the economic principle of accessing knowledge and information that is the most recent and most relevant. These texts do not deliver a detailed argumentation, subtle rhetorical or stylistic aspects, a historical overview of the problem nor any possible solution. The reading process is fast and efficient. The text gives a synthesis of the problem that is clear and up to date. The logic of books is replaced with the logic of dictionaries, user manuals, encyclopedias, vocabulary as well as short and simple texts that provide a synthesis and are read quickly. These types of short and simple texts are successful when published in collections of essays and printed as paperbacks[3]. This form of reading satisfies the need for immediate information when carrying out research. It must be completed quickly and does not allow the reader to take a long time to read and understand the text.

For example, stated in [BAR 01], this new form of reading lacks virtue but is now becoming more commonplace. Traditional reading processes that were extremely demanding could be classed as philological and hermeneutic processes. This form of reading preserved some of its biblical origins, such as difficulties encountered during the reading process and when dealing with obscurity within the text. In the traditional form of reading, these difficulties need to be mastered by the reader who must make the effort to access the transcendental sense within the text. Today, a pragmatic form of reading is taking place during which the reader chooses passages of the texts that meet his/her needs. Unaware of the traditional concept which pays respect towards literature, the new type of reader moves away from making a detailed and academic analysis of a text. This traditional form of text analysis often led to a certain sterility as texts are treated as "toolboxes". Deleuze and Foucault asked that their texts be used as toolboxes [FOU 04].

---

3 For further information on the current success of libraries see [CRI 98].

5.7. The sociological significance of different reading processes

The intelligibility of the development within reading processes also reflects sociological aspects. The significance of these two systems of reading processes becomes very interesting when they are compared to each other on the basis of the social background to which they correspond. After doing so, the political significance of changes in reading and writing processes and the usage of both become apparent. The development of public libraries is a substantial part of a democracy. What role will the digitization of documents play?

5.8. Does the "library of democracy" exist?

Robert Damien was the first to use this expression when analyzing philosophical problems of libraries. This analysis broadened the horizons of a large number of librarians who engaged with political texts and the question of how to organize a library.

In *Bibliothèque et Etat* the author shows how the concept of a public library, which is defended by Gabriel Naudé who describes it as eminently political, represents the philosophy of enlightenment by being a pillar of democracy within society[4]. *L'Avis pour dresser une bibliothèque* by Naudé (1627) represents social changes that led to the creation of a public space of constitution within a democracy. According to Damien, before establishing this space the world was strongly influenced by the Bible and libraries could only be accessed by very few privileged people. Their existence therefore remained obscure. One single book, the Bible, used to be of sacred authority. The legitimacy of all other books therefore needed to be based on the Bible. This new paradigm stated that a library "is a public exposition of knowledge that is available to everybody. This space is neither exclusive nor forbidden to some; it is a universal space of intelligibility that potentially strengthens the rationality of every human". Furthermore "democracy and the library [become] inseparable[5] on a philosophical as well as on a political [level]". Damien shows two different types of libraries and the two different corresponding political orders. The old hierarchically structured society of the "Ancien régime" corresponds to a library of the Baroque period that was characterized by "the confinement of books" and the "control of their usage as well as their meaning". The library is a monument of the obscure where the relationship to written works reveals holiness – "damaging a book [means] to insult God". Democratic societies coincided with the opening of libraries as "places where [all printed texts] can be accessed by the public". This form of

---

4 [DAM 95] the following quotes are extracts from the supplementary article [DAM 00].
5 [DAM 00, p. 35-36] ; [DAM 95, p. 16 and 29].

library is dominated by "functional instrumentality" and "trivializes the space of books[6]".

The undefined increase in the number of publications that appeared within this type of library ended the claim of the "Ancien Régime" of the existence of a single truth. The world of books experienced the downfall of old hierarchies. All books were put into libraries and functioned on the same level. All of them had a legitimate claim to truth and were in competition with each other for public recognition. Libraries became a public space and every individual could use them free from censorship and gain an overview of all printed texts to independently judge the authorities' discourse with their own enlightened intelligence. Relative, rational and recounted discourse between human beings takes place in a library free from all hierarchies infinitely open by its pluralist approach. "(…) Is the model of a coherent pluralist library at last free from the capital and therefore the source of a republican democracy[7]?"

In 1908, nearly three centuries later, Eugène Morel illustrated what this republican model of a library had turned into. In the style of a pamphlet, Morel described the "museum-library" which was outdated, dissuasive and predominant during Morel's time. With a lot of imagination he advocated a "free library" whose description seems fairly modern. He stated that "libraries will no longer be 'monuments' but agencies", "public services [in the field of] reading", "stack-rooms for public enquiries", in which the "privileged instrument" will be the "bibliographic office[8]".

This analysis is coherent when it comes to republican values. However, on a philosophical level it remains rather naive as it only considers the philosophy of enlightenment. This philosophy ignores an essential dimension without which this model remains unrealistic. There is indeed a discrepancy between the public access to knowledge and the appropriation of the public. Today, a network of public libraries that are modern and user-friendly is covering France. This network would have clearly satisfied Morel's aspirations. Furthermore, the emergence of the Internet as a digital distribution channel has led to the current situation in which these libraries are being abandoned by the public to an extent that has never been experienced before. The democratization of knowledge and culture is therefore in doubt.

---

6 [DAM 95, p. 163, 153 and 156].
7 [DAM 00, p. 36].
8 [DAM 00, p. 40].

5.9. Access and usage

The type of library mentioned above is an objective memory of intellectual work that represents the human spirit. Within libraries everybody can consult this heritage and make it come alive by reading and interpreting it. It enables people to move from written thoughts to a discussion between humans. In *Phaedra*, Plato draws attention to the gap between living thoughts and discourse that took place between a master and a student and the written discourse that should replicate this dialog. Plato criticizes those who believe that a text is "more of a reminder for those who are already familiar with the ideas discussed in the text". This criticism might appear excessive. However, we have to bear in mind that a discourse, once it has been written down, is neither able to "defend itself nor to just slip past the reader's attention when it comes to one specific idea" [PLA 00]. Written texts have their particular difficulty to them. They need to be read in order to access the information that is provided within the text. This reading process needs to be carried out by the reader. The reader's task consists of using the written text in order to understand and go back to the living discourse. According to Alain, the reader needs to "inflict his/her own thoughts on another person's way of thinking" [ALA 86].

Accessing the written form of a text is therefore not enough to ensure the transmission of the text's spirit. Immediate understanding of a text cannot be guaranteed but is based on a learning process. The necessary skills ranging from basic to very complex must be acquired. This process usually takes place at school. The access to digital information, which is based on electronic and therefore new supports of written texts, is changing these traditional skills. This task consists of bringing written texts back to life, finding the spirit behind the letters and reviving it, changing what has become objective back to a subjective view. Understanding and interpreting a text is by no means an easy task.

The opinion of those who are enthusiastic about the fact that access to documents has become easier due to digital distribution is therefore not entirely convincing as the reader needs to be able to apply the skills mentioned above to the documents in order to use them and therefore benefit from this improved access. Due to changes such as new technical possibilities, and even though the public is abandoning the library, these new developments will increase the efficiency of the library as a democratic model. Technical possibilities on their own will not lead anywhere as cognitive and cultural factors also have to be taken into account when it comes to the transmission of knowledge.

5.10. Tocqueville – a sociological model of democracy

Tocqueville defined the sociological concept of democracy. An approach that is restricted to the library cannot offer this type of definition. The Tocquevillian analysis on the way opinions are formed within democracy, combined with Martine Poulain's [POU 02] text that links it to libraries, allows for the creation of a more accurate concept.

Tocqueville's analysis mainly proves the transvaluation that marks the transition process between aristocratic to democratic societies[9]. The concepts of hierarchy and community are replaced by equality and the individual. The individual is the central axis of democracies, while equality between those individuals is close to the heart of a democracy. Tocqueville shows how these new values have an effect on the disposition of political and social spirits and the way in which things are judged and evaluated. The principle of individual autonomy is expressed in opinions and judgments. In democratic societies individuals are inclined to "search on their own and within themselves for the reason behind things". All form of authority has lost its legitimacy. There are no superior individuals to such an extent that another person could not question their judgment. Citizens therefore "always come back to their own reason as it is the most visible source that is also closest to truth". They have lost "the taste for believing in other people no matter who they are. People are isolating themselves from their surroundings and claim to judge the world based on their own reason."

According to Tocqueville this is a social aspect, a "habit of our spirit" and not simply a natural predisposition. This new "method in philosophy" where everybody "uses their own reason" could only appear "due to the fact that conditions (…) have become equal and humans have lost many of their differences". This "independence of the spirit" is a historical effect that is based on the irrepressible trend moving towards the equality of individuals.

This analysis strongly supports the "model of a democratic library". Also focusing on this topic M. Poulain draws her comparison between American and French libraries. According to M. Poulain, American libraries are there to "inform more than to provide training. Providing information is more important than the actual corpus. Skills are transmitted rather than knowledge. The same applies for the transmission of facts that is more important than the transmission of knowledge". French libraries, on the other hand, were traditionally highly respected institutions and their organization was handled with great care. Their main focus lay on training and the transmission of culture before their focus shifted towards a more Americanized concept. American libraries, as they provide information and

---

9 All quotes are taken from Chapters 1 and 2 in [TOC 81].

immediately react to the public's needs, are better adapted to the democratic spirit than French libraries [POU 02, p. 72].

Tocqueville's analysis also tempers the idyllic image of equal and cultivated individuals whose thoughts and the society they live in are based on reason and the independence of their spirit. Tocqueville believes that every society has a dogmatic feature that determines its values. These values cannot be eliminated. No individual would be capable of forming all of their opinions on their own and "following one's own truth". "Personal independence can vary in strength but it cannot be without any limits. The question therefore is not if an intellectual authority has existed during democratic eras, but where it is situated and how it acts". If we, however, try to imagine that individuals are entirely independent of one another and no "common value" exists, it becomes evident that a society such as this would not last. Shared ideas are necessary to create a society. Individuals are unable to obtain those values on their own. "It is therefore always necessary that a form of authority is formed under intellectual or moral aspects. (…) In order for an effective creation of a society, and even more so the flourishing of this society, to take place, the citizens' spirits always have to remain united on some principal ideas. These ideas might only be used so that every person compares their own ideas to them and does not simply rely on them without reflecting or comparing them."

In democracies it is not the opinion of a single person or of a group of people that creates values, but the opinion that is shared by the majority of all citizens. This is the "opinion that rules the world". Public opinion possesses "a singular power. Aristocratic regimes cannot even grasp the idea behind it. Its values are not persuasive but are imposed on everybody and become part of every single citizen's soul because of the immense pressure the common spirit puts onto the individual's intelligence". The idyllic image mentioned above is therefore drastically changing. The image of equal individuals and the independence of their spirits is taken over by an image of a regime of "isolated individuals without any defenses against actions of the masses". Pride and autonomy are replaced by social conformism[10].

Damien's approach, "the model of a democratic library", ignores the impact public opinion has on democracy. According to Damien, opinions are formed in terms of social freedom, i.e. freedom of speech and the free expression of opinions. This ideal situation is based on bibliographical plurality that is open to discussion within a group of independently thinking individuals. This situation is very similar to the unlikely event of an "ideal communication situation" as Habermas describes it. When analyzing this model, sociological and cultural aspects need to be added to

---

10 "Citizens that want to judge for themselves are most likely those who follow the majority's opinion." In this quote Descombes insists on the sociological dimension of Tocqueville's analysis [DES 04, p. 369].

the public's rights. The library on its own will not become a "matrix of reason" [DAM 95, p. 127]. Neither technical possibilities in the field of distribution and access to information, nor the right to access information and freedom of speech will be sufficient to guarantee the democratization of culture. They are necessary but not enough on their own. The library can only function as a paradigm of a public democratic space if there is the common tradition of using the library in this way. Equal access to libraries does not determine the way in which they are used. Users remain prisoners of their own or society's ideological concepts and therefore due to sociological predetermination miss out on real access to knowledge and culture.

There is no single individual ready to absorb all of the never-ending treasures of heritage that can be found within a library. Nobody would actually be capable of finding their way around the labyrinth of every text the human spirit ever produced. Normative criteria of orientation and evaluation, as well as principles of which text types should be avoided, are given to the user by the society that they belong to. Without a normative orientation that society brings about, the user would be entirely lost within the universe of all texts that have ever been produced.

The approbation of written culture, whether it is printed or digital, depends on "common sense", i.e. the sense that is attributed to written texts by society as a whole and not only to the sense the individual attributes to it. Not only society, but the individual, represents itself within written culture. The individual is also subject to a common opinion, i.e. the anonymous opinion of the majority. This opinion and the dispositions that underlie it are in the widest sense produced by education. This certainly includes education in school but also, and even more so, the social context in which audiovisual media, sometimes even digitized media, play such a powerful role today.

5.11. The library's devices and the disposition of the public

The way in which a library is organized, no matter how ingeniously this is done, never simply imposes a structure on the documents on offer. No communication policy is capable of determining the readers' demand[11]. When only considering the immediate demand, we might discover that those in favor of the Internet and who insist on the improved communication and the universal availability of publication, are wrong. Democratic processes within knowledge and cultural creativity do not only depend on the fluidity and the intensity of communication as the advocates of the Internet like to believe. In reality, progress in this field depends on the cognitive processing of information and not on the way it can be accessed. Furthermore, the

---

11 [BAT 03] mentions the hubris for the possibilities the library offers for the reformation of culture and society, p. 120.

value that a certain society attributes to knowledge and culture plays an important role.

How can the tools and offer of a library respond to the dispositions of the public? This, of course, depends on the public and at this point attention should therefore be drawn to perhaps the biggest problem in the traditional organization of a library. A library does not provide guidance on the quality of the texts in their collections, nor information on how difficult a text is, or its level of intellectual importance. The presentation of a library's collection does not follow any kind of hierarchy. However, their corpus has the function of collecting and allowing access to texts. Furthermore, libraries have the function of gathering together all classics of the same discipline[12]. The traditional concept of libraries states that their role is to provide free and universal access to information. Libraries ought to be free of any value judgment. In reality, the ways in which librarians make information accessible, i.e. the way in which this information is organized, structured and presented as well as classification in a catalog and indexation, are never entirely neutral. Librarians are aware of this fact. A very good example are the collections and their classification that date back to the 19th century (Dewey and UDC) which show the state of knowledge at the time and this era's outlook on the world with all its values and prejudices. Dewey, as we know, is entirely focused on the USA.

A classification in the form of an encyclopedia always has a substantial impact on the presentation of knowledge. All forms of organization lead to a certain hierarchy, and choosing certain disciplines necessarily implies value judgments on these disciplines and their objects. It is impossible for the vocabulary of humanities to be as neutral as that used in science. It depends on the particular cultural context, value judgments on the existence of human beings, and the way in which these aspects are defined in different societies. Attempts to make a thesaurus more objective are sometimes ridiculous and shock the readers or make them laugh. Here are some examples from the RAMEAU language, "Indians: first contact with Europeans" is used to describe the Spanish landing and colonization of Latin America. As another example, the term "socially handicapped" is used to describe individuals that suffer from social exclusion.

Criticism inspired by Foucault on the "occidental ideology of reading" was expressed by Armando Petrucci. "Hierarchies and values (...) which are introduced mechanically into the structures that are used for preservation and usage (...) have become a source of authority within themselves. Ordinary readers will not question this judgment which will in the end be integrated into general public opinion" [PET 97]. This challenge can only be mastered if libraries do not try to give the

---

12 For more details on this point, see [BEG 00].

impression that all classification can be abandoned. This idea would lead to a friendly anarchy within the library and could contribute to the creation of an alternative canon. The "aversion process" leads to a library that "pluralizes its sources of authority, multiplies the perspectives, increases its references and allows for free movement". Damien gives an excellent description that is very likely to be carried out and at the same time a very well founded argument [DAM 00, p. 36]. Indeed, when facing the fact that libraries are abandoned and an "overproduction of written texts" (Baudrillard) is currently taking place, the reader easily turns to a simple solution that is advertised by Internet search engines. How can the reader not be skeptical, and even disgusted with the astronomic dimensions of memory and the strong acceleration in the number of articles that are published? This infinite plurality does not lead to the questioning of knowledge within democracies, but rather paves the way for "the denunciation of bibliophobia" that Damien mentions [DAM 00, p. 36].

5.12. Libraries are facing a cultural crisis

The conditions of democratic access to knowledge and culture within its entire plurality, which include allowing access to controversial material, is part of education and transmits knowledge as well as skills. As previously mentioned, there are different levels of reading and writing. Mastering texts therefore requires a number of skills that are transmitted mainly via specific education in the field. It is, however, still possible to actively engage with the content and understand the information that is offered. Apart from being able to read, two other skills – understanding and interpreting a text – are also required. The difficulty of these last two skills lies in the fact that spirit and knowledge are required which have to be adapted to the time period during which the text was written. The capacity of choosing knowledge and information that is relevant to the text is particularly important. Classification, indexation and bibliographies are the pillars of a library and should provide a framework, i.e. guidelines for the reading process. They are, however, entirely insufficient on their own. It is therefore necessary to develop cultural knowledge, i.e. general knowledge in order to develop the skill that enables the reader to distinguish the importance of information in different texts. A value judgment, or in other words the skill to establish a certain hierarchy of ideas, is the basis of orientation and order within a text. Orientation within a text includes the concept of discriminating against certain ideas. This necessary skill does not agree with the democratic spirit. Individuals find it very difficult to introduce this necessary hierarchy as they live in an egalitarian society[13].

According to M. Gauchet [GAU 02] the society of individuals has a tendency to form "individuals that depend on the pretext of autonomy." The values that made a

---

13 This point is covered in detail in [DUM 83].

republican education possible are currently experiencing a crisis. It is difficult to explain and justify how a library can still function as an "institution of a political matrix that fosters the citizen's judgment[14]". Concepts such as the "bibliographic office" and a "public reading service[15]" that Morel suggests as the key point of a library seem much better adapted to the current situation. A good example of this new trend is university libraries. Their bibliographic offering is currently diminishing as there is a lack of demand. The offer of training in bibliographic research cannot be maintained if it remains for the public to choose whether they would like to participate[16]. Given the decreasing usage of books and the increasing use of the Internet, without any further thought on how to use the Internet, the democratization of culture does not seem to be taking place, even though an ever increasing number of students are enrolling in university.

The crucial question of values within our society is attached to knowledge and culture. What is the public's disposition towards cultural heritage and printed heritage? What place will the public attribute to this heritage and how will they define it? What is the public's image of reading culture and what value does it attribute to written expression? What are the expectations towards libraries and what significance do digital applications have[17]?

For the library's treasures and tools to be used effectively, the taste for knowledge and culture as a core value needs to be transmitted to future generations. Otherwise, the "occidental canon" of written texts will be "questioned". We are currently experiencing a "crisis of institutionalized structures and ideologies that until now have 'supported the old order of reading processes[18]'". Cultural heritage, "as norms and values, as forms of socialization and culture, as a social historical background of individuals, as a significant relationship of the community with itself and its members, as a link to time and its own written works, is profoundly questioned (…)" [CAS 78].

---

14 [DAM 00, p. 37].

15 [DAM 00, p. 39 and 40].

16 The organization of a library does allow the general public easier access to information. However, the general public still needs to familiarize itself with the way in which the library is organized. The reader needs to understand the basic outline of classification and needs to be able to deal with catalogs and databases. The library's role therefore does not only consist of the organization of collections according to precise norms as well as the production of catalogs and indexation, but also in the training of the reader when it comes to their orientation within the library.

17 [FLI 01] showed different meanings of electronic networks that have successively been integrated and make up the different layers of ideology.

18 [PET 97, p. 423]. "The current situation presents the symptoms of dissolution that is expressed by the organization of reading processes within the occidental culture of reading."

5.13. Conclusion

Neither technical and instrumental possibilities that emerged with digitization, nor the tools of traditional libraries on their own can lead to the creation of an original culture. In other words, all of them are required to contribute to the democratic distribution of cultural heritage that is an essential step in the creation of an original culture. Digital applications as they are used today are not being developed in this direction. They favor fast access to information and insist on unilateral communication to the detriment of transmitting the idea that further reflection is required.

Three conditions are required when using digital applications and the Internet for progress in the democratization process of knowledge. These three conditions also allow for strategic training in order to improve the citizen's judgment; something that Damien hopes for. Digital libraries first of all need to be created, since the dominant logic of online search engines leads to the current mediocre level of directories. According to Kuny and Cleveland, "a digital library would include all the processes and services – collection development and management, subject analysis, index creation, reference work and preservation – that are the backbone and nervous system of contemporary libraries" [KUN 98]. Training in bibliographic research needs to be promoted and the methodology of documentation needs to be integrated into the curriculum of all forms of education within school[19]. Last but not least, the expectations on all different levels of education need to be re-evaluated and the focus needs to be shifted towards a broad general knowledge that will allow for guidance within the never-ending universe of written texts. The most difficult, and therefore also the most important task, is teaching cultural heritage in an era where subjectivity is regarded as the ultimate goal. Teaching people how to appreciate the infinite debt to the past without being conservative or putting too much emphasis on the past is of utmost importance. Finding an adequate compromise between the hierarchically structured traditional culture and the individualist logic of the digital culture, i.e. between the republican model of transmitting knowledge and culture and the free market of information, is of utmost importance.

5.14. Bibliography

[ALA 86] ALAIN, *Propos sur l'éducation*, PUF, p. 96-102, 1986.

[ARO 91] ARON R., *Introduction à la philosophie de l'histoire*, Gallimard, 1991.

[BAR 01] BARBIER-BOUVET J.F., "Internet, lecture et culture de flux", *Esprit*, December 2001.

19 For information on bibliographic research and university libraries please see [FRE 02].

[BAT 03] BATTLE M., *Library: An Unquiet History*, Norton & Company, New York, 2003.

[BEG 00] BÉGUET B., "Des corpus dans un plan de classement", *BBF*, no. 3, 2000.

[CAS 78] CASTORIADIS C., "Transformation sociale et création culturelle", *Sociologie et sociétés*, Montréal, December 1978.

[CAV 97] CAVALLO G., CHARTIER R., *Histoire de la lecture dans le monde occidental*, Seuil, p. 8, 1997.

[CRI 98] CRITIQUE, *Dicomania ou la folie des dictionnaires*, no. 608-609, February 1998.

[DAM 95] DAMIEN R., *Bibliothèque et Etat*, PUF, 1995.

[DAM 00] DAMIEN R., "Procès et défense d'un modèle bibliothécaire de la démocratie", *BBF*, vol. 45, no. 5, 2000.

[DES 96] DESCOMBES V., *Les institutions du sens*, Minuit, 1996.

[DES 04] DESCOMBES V., *Le complément de sujet*, chap. XLVI, p. 369, Gallimard, 2004.

[DUM 83] DUMONT L., *Essais sur l'individualisme*, Seuil, 1983.

[EIN 91] EISENSTEIN E., *La révolution de l'imprimé à l'aube de l'Europe moderne*, La Découverte, Paris, 1991.

[FLI 01] FLICHY P., *L'imaginaire d'Internet*, La Découverte, 2001.

[FOU 04] FOUCAULT, "Entretien avec Roger-Pol Droit", *Le Point*, July 2004.

[FRE 02] FRESSARD O., "Paris-8, pionnière en formation des usagers", *Bibliothèques 5/6*, December 2002.

[GAU 02] GAUCHET M., *La démocratie contre elle-même*, Gallimard, 2002.

[GOO 96] GOODY J., *La raison graphique*, Minuit, Paris, 1996.

[IOR 00] D'IORIO P. (ED.), *HyperNietzsche*, PUF, 2000.

[JAC 99] JACQUESSON A., RIVIER A., *Bibliothèques et documents numériques*, Cercle de la librairie, p. 329, 1999.

[JEA 00] JEANNERET Y., *Y a-t-il (vraiment) des technologies de l'information ?* Presses du Septentrion, 2000.

[JEA 01] JEANNERET Y., *Nouveaux outils, nouvelles écritures, nouvelles lectures*, *Spirale*, no. 28, October 2001.

[JOH 94] JOHANNOT Y., *Tourner la page : livres, rites et symboles*, J. Millon, 1994.

[KEN 91] MCKENZIE D.F., AMFREVILLE M., *La bibliographie et la sociologie des textes*, Cercle de la Librairie, Paris, 1991.

[KUN 98] KUNY, "The Digital Libraries: myths and challenges", *IFLA, Journal*, vol. 24, no. 2, 1998.

[LAB 79] LABARRIÈRE J.F., *La Phénoménologie de l'esprit de Hegel : introduction à une lecture*, Aubier-Montaigne, p. 25, 1979.

[LEF 72] LEFORT C., *Machiavel : le travail de l'œuvre*, Gallimard, 1972.

[PAR 91] PARROCHIA D., "La difficulté de classer", *Mathématiques et existence*, Champ Vallon, 1991.

[PET 97] PETRUCCI A., "Lire pour lire : un avenir pour la lecture", *Histoire de la lecture dans le monde occidental*, Seuil, p. 401-425, 1997.

[PLA 00] PLATON, *Phèdre*, Flammarion, p. 178 and 183, 2000.

[POU 02] POULAIN M., "Retourner à Tocqueville pour comprendre l'histoire comparée des bibliothèques américaines et françaises au XIXe siècle", *BBF*, no. 5, 2002.

[RIC 66] RICHAUDEAU F., *Cours de lecture rapide*, Centre d'étude et de promotion de la lecture, 1966.

[STI 91] STIEGLER B., "Machines à lire", *La Bibliothèque*, Autrement, série Mutations, no. 121, April 1991.

[TOC 81] TOCQUEVILLE A., *De la démocratie en Amérique*, II, Flammarion, 1981.

[VAN 99] VANDENDORPE C., *Du papyrus à l'hypertexte : essai sur les mutations du texte et de la lecture*, p. 41-50, la Découverte, 1999.

[VIR 93] VIRBEL J., "La lecture assistée par ordinateur et la station de lecture de la Bibliothèque de France", *Banques de données littéraires, comparatistes et francophones*, Pulim, 1993.

[ZYS 99] ZYSBERG A. *et al.*, "La lecture assistée par ordinateur", *Le Débat*, no. 86, September/October 1999.

Chapter 6

Accessing Library Catalogs in the Age of Digital Libraries and Search Engines: Gaps, Disruptions and Transformation?

I am only here as one member amongst many others. The "secret order" I am part of will be discussed later on in the text. This article strays from the topic in two ways: I will be talking about libraries which are neither digitized nor specialized.

I will focus on libraries that are not digitized because even though the dematerialization of information is thriving, a large number of libraries still provide their readers with printed documents. I would like to emphasize that it is not copyright regulations which force them to do so.

I will focus on non-specialized libraries as the secret order I am part of, and which is not meant to last, currently only recruits its members in public, municipal and regional libraries.

6.1. Prehistory

First of all, the era before IT will be covered in this chapter. Libraries, just like societies, took a long time to leave the prehistoric state behind them. Today, some of them are still stuck in this period of time.

6.1.1. *Secondary information*

To manage their stock, libraries needed to create information on information. Today this information on information is referred to as metadata.

Other than describing the standards for bibliographies and languages of documentation, it could also be useful to take a look at the support that was used for this type of information.

6.1.1.1. *Printed catalogs*

The advantage of printed catalogs was that they allowed for remote access. This is why the majority of libraries continued printing their catalogs throughout the 20th century[1]. The National Library of France (BNF) continued to print their catalogs until IT took over.

The weakness of printed catalogs is the difficulty in updating them. In order to do so, either a new version of the catalog needs to be printed or supplements have to be published. The production of supplements, however, makes research more difficult.

6.1.1.2. *Catalogs on microfiche*

The problem of updating catalogs was solved by used microfiche. Inserting new information became one of the main activities for professional librarians. This task was so important that it became subject to professional examinations or competitive entrance exams.

On the other hand, the possibility of accessing the catalog from a distance was lost.

The catalog could only be accessed by one person at a time. When someone was consulting a slide valve their vertical usage of it blocked an entire column.

Furthermore, several microfiches had to be created for different types of access or modes of search, e.g. one for the author, one for the title and one for the topic.

---

1 In the 1980s the Conseil général (council of a French department) of Hauts-de-Seine sent out a questionnaire to its local libraries. One of the questions asked was: "When was the latest edition of your catalog published?" This question was, apart from a spelling mistake which was added in the 1980s, exactly the same question which had already been asked by the prefecture of the Seine in the 1930s.

The creation of multiple microfiches was a heavy strain on the successful management of the system[2].

6.1.1.3. *Strengths and weaknesses of non-digitized catalogs*

The material used for the production of catalogs before the emergence of IT has left a mark on practices in the profession. Conciseness is a good example as it has become an imperative:

– *descriptive information needs to be very concise.* Limited by the material support of a catalog, i.e. a microfiche or a printed page of a catalog on which a certain number of references needed to be published, the librarian's ideal support is the ISBD[3] and a format that could be referred to as "points-spaces-dashes";

– *concise access*. With the limited number of search entries[4] a specialized jargon was created in documentation. It summarized the content in so-called subject headings or key words.

However, these achievements also have their downfalls.

– the standardized and concise description of documents is not always easily understood by the library's users;

– limited access and the rigidity of the jargon used in documentation do not make any sense when the limits imposed by microfiche or paper no longer exist.

One major advantage of printed catalogs, or those published on microfiche, is vague research. The user simply leafs through the catalog, his/her research is vague as the catalog is explored within the semantic neighborhood and based on orthographical proximity.

This intimate experience of research can be understood by those who experienced the era before the emergence of IT. This type of research process could be described by the term *serendipity*[5].

---

2 On the other hand, the Minigraph created duplicates using stencils, and later typewriters with an integrated memory allowed for the production of several microfiches by hitting one single key.

3 ISBD: *international standard bibliographic description*.

4 The indexation standards brought, for example, the number of authors of a book down to three. If a book was written by more than three people this was indicated by the nice little sentence "anonymous due to too many authors".

5 *Serendipity* is a word that was invented in 1754 by the English philosopher Sir Horatio Walpole. It describes the ability of some people to find the required information intuitively without spending much time on research. The word stems from a novel written at that time called *The Three Princes of Serendib*. Chance leads them to the right solution in difficult situations [ERT 03].

6.1.2. *What about access to documents?*

For a very long time the user was unable to directly access documents. Consulting the catalog first was an obligatory procedure.

Free access to bookshelves changed the role of the catalog which, from that point onwards, was only regarded as a complementary tool. Very often users go directly to the shelves.

6.2. The age of OPAC

The abbreviation OPAC (*Online Public Access Catalogue*) is used all over the world and describes a digital catalog search interface within libraries. "Consult the OPAC" is one of the sentences we often hear when members of staff of certain libraries talk to a user.

Is use of the OPAC entirely understandable to the user[6]?

Figure 6.1. *People consulting the OPAC at the public library in Shanghai. Photo by Dominique Lahary, July 2000*

---

6 To find online catalogs of French libraries the following directories may be used: French Ministry of Culture: http://www.culture.fr/BibliothequesMediatheques/c371, Sibel: http://sibel.enssib.fr/index.php ?m=c&c=418, association of directors working at regional libraries (ADBDP): http://www.adbdp.asso.fr/lesbdp/cataloguesbdp.htm, association of directors working at local and district libraries of larger cities in France (ADBGV): http://www.adbgv.asso.fr/index.php?page=opacsweb.

6.2.1. *A high level of uniformity*

The standardization of data already existed in the form of catalogs on microfiche (e.g. MARC[7] formats). The same standardization process was also used in the OPAC.

At the beginning, the OPAC was subject to different interfaces which changed from entering characters to clicking on icons. It has finally been transformed into a web application. The OPAC can be consulted online, or in most cases, only from within a library.

The OPAC unites different structures:

– the following succession of different types of search entries will appear on screen:

- a search form, often subdivided into two different versions: a basic search and a detailed search,

- a list of terms, this stage is not required in certain types of search, when looking for a specific title for example,

- a list of abbreviated descriptions of the records,

- descriptions of the records as a full text;

– when it comes to fields for search entries the trinity of author/title/subject is dominant.

6.2.2. *How to access documents according to their content*

The main focus of this article lies in searching documents according to their content.

In the case of printed catalogs or catalogs on microfiche, access is organized by librarians who use a specialized jargon. As a result, we encounter a mixture of too much and very little specialized vocabulary;

– the large amount of vocabulary and the complexity of it can be structured in a certain hierarchy as has been done in RAMEAU[8];

---

7 MARC: *machine readable cataloging* (a type of format that emerged in the USA in the 1960s).
8 RAMEAU: directory that resembles an encyclopedia published in alphabetical order, it is widely used in France amongst suppliers of bibliographic information such as the National Library of France, Electre and SUDOC (a bibliographic network of universities and other institutions of higher education) as well as by a several public libraries.

– the lack of vocabulary can be encountered when it comes to the jargon used by the people who are searching the system.

6.2.3. *Too many or no hits at all – a choice must be made*

The system is laid out in a way that suggests that the main aim of librarians is to avoid too many hits when carrying out a search[9]. The result is that the user often does not obtain any search results at all[10].

The librarian's obsession with obtaining relevant results requires very precise search entries. However, the user is unable to use the librarian's jargon nor is he/she aware of the way in which this jargon is used in the individual catalog.

In reality, a higher number of hits, even if they are not entirely relevant, does not necessarily mean that there are too many. Online search engines present an extremely large number of hits. However, these hits can be sorted using different techniques. Competition between different search engine operators is mainly based on the criteria mentioned above. In this case, a high number of hits will not be chaotic at all but could actually be seen as a perfectly arranged panorama picture that allows for an overview of all possible data.

6.2.4. *Some progress is being made*

Digital catalogs, of course, have certain advantages over printed catalogs or those published on microfiche:

– search criteria can be linked to each other. The Boolean operator leaves the user the choice of and/or/not;

– all, or some, fields of the bibliographic descriptions of the records can be searched at once.

We might also refer to these new trends as linguistic processing (e.g. orthographic proximity, grammatical inflections).

---

9 Any hit that might not be relevant in a search of documents [VOC 93].

10 When carrying out a search the phenomenon of very few hits represents the problem that documents which are relevant to the search entry are not selected as their descriptions are not linked to the search entry.

6.2.5. *Disadvantages and features the system lacks*

Even if it is sometimes possible to take a look at the index, it is impossible to leaf through the catalog.

This is frustrating for users as they cannot choose between documents. Instead, they have to choose between different concepts such as the index of authors, the index of topics etc. This is how the user finally gets lost in the jungle of jargon that is used in the field of documentation.

Precise research does not take the principle of serendipity into account. The German aphorist Georg Christoph Lichtenberg (1742-1799) illustrates this fact in one of his *Sudelbücher:* "Many people, probably the majority, need to know what they are looking for before being able to find it."

6.2.6. *Are catalogs actually used by the public?*

Library users are very rarely also users of catalogs. A study known as "*sortie de bibliothèque*" which was carried out in 18 libraries in the east of the French region Val-d'Oise in 2003 confirmed this statement.

People who answered yes to the question "have you found what you were looking for?" were also asked how they managed to find the required information. There were several possible answers to the second question.

4% of the users questioned in rural areas of France and 13% of users questioned in more urban areas stated that they used the catalog to find the required information, 16% (rural areas) and 22% (urban areas) stated they found the desired document by chance, 20% (rural areas) and 24% (urban areas) obtained the required documents from a display shelf. The majority, however, 69% in rural and 70% in urban areas, found the required document by leafing through the stack shelves.
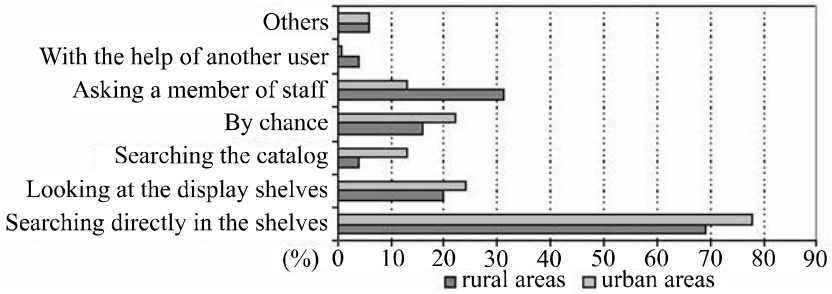
Figure 6.2. *Searching for documents: multiple answers*

If more than 55% of all users combine several methods, it is interesting that in this case only 0.5% (rural areas) or 2.5% (urban areas) use the catalog. 7% (rural areas) or 4.5% (urban areas) ask a member of staff while 31% of all users in rural areas and 37% of all users in urban areas search directly in the stack shelves.
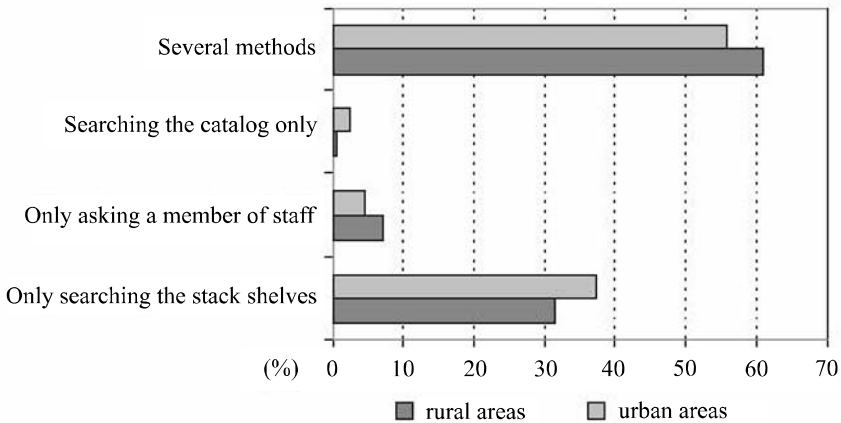
Figure 6.3. *Searching for documents: combining different methods or only using a single method to find a document*

This study confirms that catalogs, whether they are digitized or not, are only consulted by a very small percentage of users that visit the library.

Catalogs are useful for people who are not in the library but would like to obtain information on the library's resources remotely.

6.3. The secret order

6.3.1. *Libraries must now imitate search engines which so successfully imitated them in the first place*

The website of a library could easily follow the same pattern as a physical library, built for receiving users:

– at the top of the page there would be links to information that give a general overview and descriptions of the services offered. One field would allow for a "quick search". A form for an "advanced search" would require the following search entries; author, title, subject and other key words. They could possibly be linked with a Boolean logic;

– on the right hand side news, summaries of periodicals and local community or university services could be shown.

The left hand side and the center of the screen, i.e. the largest part of it, could be devoted to subject trees such as Dewey's classification system that explains how documents are linked to one specific subject area[11]. These topics enable the user to access the catalog and decide on criteria according to which the results are sorted, e.g. chronological order.

What has been described above resembles exactly the homepage of widely known search engines such as Yahoo. Only some words would need to be changed to create the library portal. Let us try to imitate those who, maybe without realizing, imitated us so well.

What is suggested here is (…) to put the focus back on catalogs, on a systematic access to documents. This type of classification had been neglected since intellectual classification systems and free access to collections were introduced. Increasing the importance of catalogs would be very efficient because all indexes are clearly labeled and separated into different divisions and classes.

Using subject trees every Internet user is familiar with does not draw the user's attention away from the actual question, but helps them to find what they are looking for without having to search for the right question to ask in the first place[12].

---

11 Subject areas are independent of a systematic classification of books that traditionally separated them into different fields, such as childhood or nature.
12 Extract from [LAH 00].

6.3.2. *The secret order's manifesto*

Based on this idea some librarians have created an informal group. They all met for a second time and accepted a manifesto that was published in November 2002[13].

We are convinced of the following facts:

– finding the required information is more important than the search process;

– not enough hits is worse than too many;

– the free access to shelves also needs to be reflected in online libraries;

– online directories and web resources are models that should be followed. Members of the secret order agree to exploit all possibilities, programs and interfaces of current and future libraries to supply the public with access to a catalog that is subdivided into subject areas.

Alain Caraco indicates the objectives of this project[14]:

Online directories are the model that should be followed during this project.

– hierarchical classification systems (an adapted version of Dewey's list whose parameters may be changed), subject trees, going from general to specific information, transversal links entitled 'also refer to' should be introduced;

– a simple search with only one field and an index enable the user to limit a search entry to only one field of the subject tree or search the entire catalog.

6.3.3. *Plea for resurgence*

Richard Roy created the website "Le butineur: expériences, références, projets en matière de facilitation et d'optimisation de l'accès aux collections des

---

13 Published on this page: http://membres.lycos.fr/vacher/profess/accesweb.htm. List of members as of 30 July 2004: Aline Calenge, regional library of Saône-et-Loire, Alain Caraco, director of Chambéry's municipal library (Savoie), Quentin Chevillon, municipal library of de Saint-Cyr-l'Ecole (Yvelines), Victoria Courtois, regional library of Val-d'Oise, Elisabeth Derderian, multimedia library 'Maison du livre, de l'image et du son' of Villeurbanne (Rhône), Michel Fauchié, multimedia library of La Roche-sur-Yon (Vendée), Sylvain Fontaine, multimedia library of des Ulis (Essonne), Régis Faivre, director of Doubs' regional library, Dominique Lahary, director of Val-d'Oise's regional library, Daniel Le Goff, director of the regional library located near Aisne, François Lemarchand, director of Agneaux's multimedia library (Manche), Pierre Louis, director of Metz's municipal library (Moselle), Marie-Christine Pascal, director of Châlon-sur-Saône's municipal library (Saône-et-Loire), Richard Roy, municipal library of Reims (Marne).
14 http://alain.caraco.free.fr/index.php?page=opac.

bibliothèques publiques[15]" (The browser: experiences, references, projects aiming at a better and easier access to collections in public libraries). Some of his questions remind us of a speech that was given at the ISKO[16] (International Society for Knowledge Organization) congress that took place in Grenoble in July 2003. This speech dealt with the systematic way in which manual collection catalogs that were kept in stack rooms were organized. The documents are presented within a classification system, e.g. the Dewey decimal classification system[17]. This systematic catalog was a supplement to the catalog stating all subjects in alphabetical order or, of course, the catalogs giving the titles and names of authors in alphabetical order. Thus, the user could leaf through them.

With users being able to directly access the shelves, the traditional classification system way replaced by a new order. Documents were arranged according to one or several classification systems. Systematic catalogs were often abandoned.

According to Richard Roy these catalogs, which had previously been abandoned, are now re-emerging due to the Internet, as they are published online. When using the Internet the user is unable to actually take a look at the stack shelves. The only way to access the library is to type in a search entry. It has become impossible to discover the collection as a whole and explore it from its larger to smaller subject areas and find hidden paths that lead to unexpected discoveries within the semantic neighborhood.

Richard Roy concludes as follows:

It is therefore necessary to add a topic-based form of presentation to the online library catalogs. This type of subject tree will enable online users to search the library without having to type in a search entry. This approach is closest to the situation of a reader wandering through the stack shelves of books.

---

15 http://site.voila.fr/le_butineur/web.htm.

16 http://site.voila.fr/le_butineur/isko2.htm.

17 The decimal classification was introduced and regularly updated by the American Melvil Dewey (1851-1931). Even though the actual rooms need to be adapted to Dewey's classification system, it is used in the majority of French public and university libraries; see [BET 98] and the article on Dewey's decimal classification system published on http://www.oclc.org/dewey/versions/webdewey/default.htm.

6.3.4. *Realizing the project*

6.3.4.1. *How members of the secret order carried out the project*
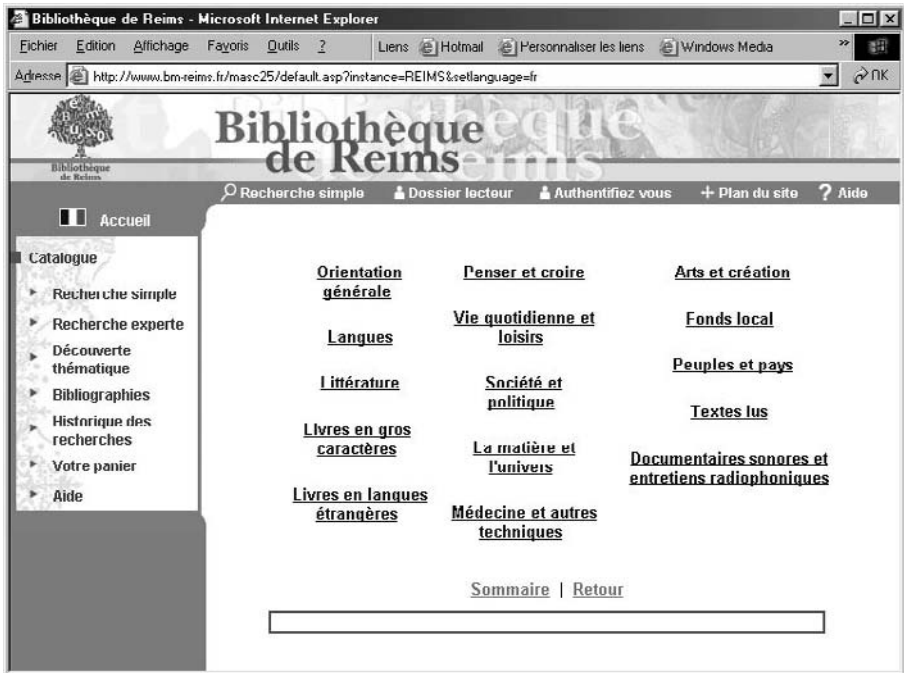
Figure 6.4. *Homepage of the municipal library in Reims showing the subject tree of the online catalog*

Members of the secret order tried to implement the aims laid out in the manifesto within their own institutions.

Alongside traditional ways of accessing their online catalog, two libraries offer a topic-based access based on Dewey's elements of classification:

– Reims municipal library has offered this type of access with the introduction of their "subject tree" that enables the user to discover the collection based on different themes[18];

– the regional library of Aisne offers an interface where on the same screen keywords can be entered or a list of subject areas can be accessed directly[19].
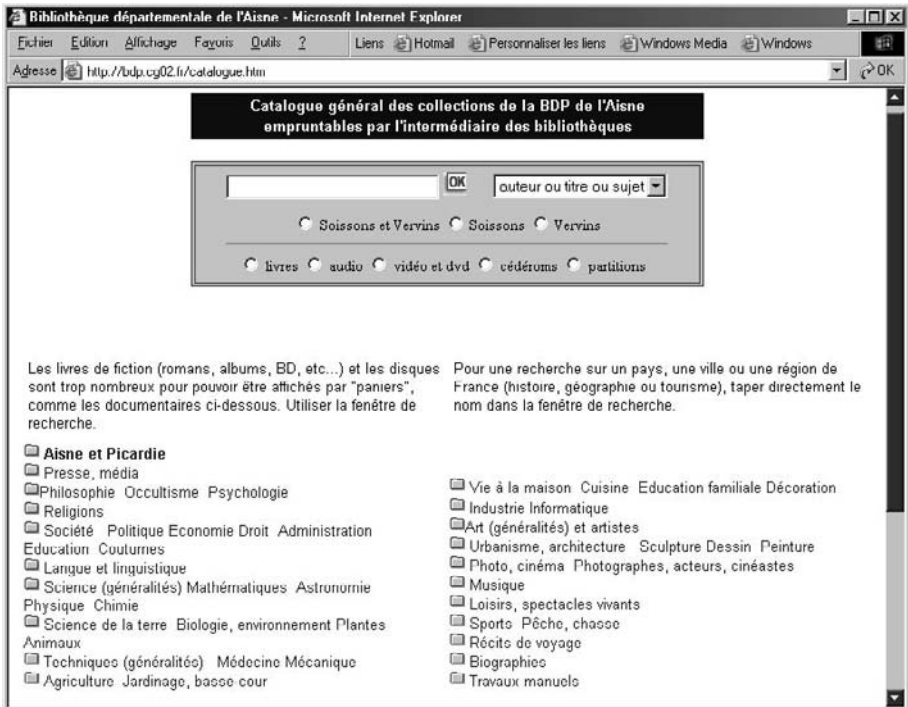
Figure 6.5. *Homepage of the online catalog of the regional library of Aisne, France*

18 http://www.bm-reims.fr/masc25/default.asp?instance=REIMS&setlanguage=fr. The content management tool Absys and the interface were created by a company called Archimed and adapted to the needs of Reims municipal library.

19 http://bdp.cg02.fr/catalogue.htm. The content management system Aloes was created by a company called Opsys who also participated in the installation of this interface.

The regional library of Val-d'Oise offers several interfaces on a topic-based search on its homepage[20]. Different topics that might be chosen are, for example, illustrated books for children, expositions, topic-based presentation of CDs and tapes for young people, presentation according to the different genre of music, books or CDs. A subject tree has been created for the first three topics mentioned above. Music is classed according to the PDCM (principles for the classification of music documents applicable to library collections[21]).

Figure 6.6. *Homepage of the regional library in Val-d'Oise. The interface shows a theme-based access to illustrated books for children*

---

20 http://www.valdoise.fr/biblio/bdvo/. The company Sirsi created the content management system Multilis and the interface web DRA-web2. This adjustable interface was adapted by Victoria Courtois who is responsible for new technologies as well as information systems and communication NTIC at the regional library of Val-d'Oise.

21 This type of classification was published in [MUS 02].
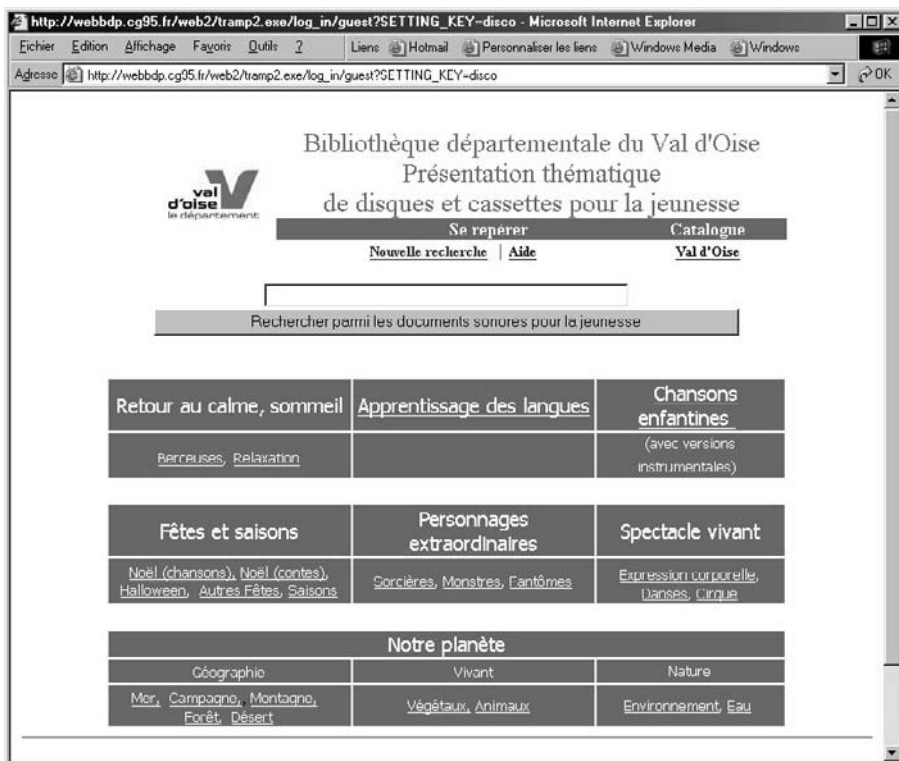
Figure 6.7. *Homepage of theme-based presentation of CDs and tapes for young people at the regional library of Val-d'Oise*

The multimedia library of Agneaux (Manche) offers a catalog that enables the user to browse through different categories and is based on the *Grumeau* thesaurus which was especially created for this task.

Last but not least, MOCCAM[22] is an open source program written by Quentin Chevignon which works as a collective virtual catalog. It provides a traditional search engine similar to Yahoo[23] that enables the user to browse through categories of Dewey's classification as well as a search based on the keywords of this classification. When entering a term MOCCAM also indicates Dewey's sections that contain one of the terms.

---

22 My personal collective catalog: http://www.bmtriel.no-ip.org/ccy2.

23 http://www.moccam.no-ip.org/ccy2/modules/xoopsfaq/index.php ?cat_id=4#q14.

6.3.4.2. *How French software companies carried out the project*

On his website Richard Roy has established and continually updates a directory of software companies[24].

After the manifesto had been published, several companies created a topic-based access to catalogs:

– the company Opsys created a program called Aloes. In addition to the regional library of Aisne, the municipal library of Dijon[25] also uses this program to provide access to themes such as literature for young people, different genres of novels and different genres of music;

– with its program *Paprika* Décalog offers two different types of topic-based access. This program is used in the municipal library of Vienne[26], where Dewey's classification is used for books, documentaries, and the PDCM standard is used for music, as well as in several French centers for culture abroad[27], where this program provides access to different fields of interest that have previously been established by the company.

Creators of open source such as Koha[28] or PMB[29] are emerging and are progressively offering an increasing number of solutions for content management systems for libraries.

6.3.4.3. *Other attempts to carry out this project*

The members of the secret order are, of course, not pretentious and do not claim to have invented the concept of creating topic-based access for digital libraries. They recognize similar applications and other interfaces used for research.

When it comes to bibliographic references that do not stem from libraries, we can quote the bibliography of the SMES (society of medieval historians working in public higher education) that provides over 10,000 descriptions of the records which are structured in a hierarchical order[30]. Many other libraries worldwide also use

---

24 Please refer to http://site.voila.fr/le_butineur/web.htm, section: Réalisations.

25 http://www.bm-dijon.fr/c103.htm.

26 http://193.251.9.79/opacweb.

27 Instituts français in Madrid (http://www.ifmadrid.com), Budapest, (http://mediatheque. inst-france.hu/Opacweb), Prague (http://ifp.paprika.net/Opacweb), Rio de Janeiro (http://medrio.paprika.net/opacweb) and New Delhi (http://firc.paprika.net/opacweb/).

28 http://fr.koha.org.

29 http://www.pizz.net.

30 http://www.mom.fr/shmes/biblio.

topic-based search tools. These can be youth interests such as the one in Auckland (New Zealand[31]) or popular subjects like in Kansas City (USA)[32].

6.3.5. *Remote access*

Members of the secret order are certainly not under the impression that they have found the ultimate solution to all problems.

On the one hand, they acknowledge catalog projects that have already been carried out abroad and whose solutions are based on the same principles.

On the other hand, they communicate via email in order to compare and comment on how the project has been carried out by different libraries or companies. They are not ignoring the downfalls of certain attempts to carry out this project. During this communication process three main challenges have been established.

6.3.5.1. *The number of hits is too high*

"Too few hits is worse than too many": this statement is certainly true but too many hits does also not represent an aim in itself. How does a user react when looking at the section "handicraft" and being faced with more than 9,000 descriptions of records that are not at all sorted according to their relevance? More generally speaking, this type of search engine necessarily selects several hundred descriptions of records for every section, (in the case of a catalog of more than (...) 50,000 documents this means that there is an average of around 100 sections with 500 descriptions of records per individual section[33])."

The problem of the number of hits is especially worrying as it creates a dilemma for those who program the interface. Increasing the number of sections would bring down the number of hits but does not improve the situation as most users abandon their search if it takes more than two clicks of the mouse to find what they are looking for.

Others, however, are ready to accept a never-ending number of mouse clicks on subject trees or online directories.

The number of users who abandon their search could be reduced if the hits were sorted according to their relevance. The library's principle of sorting the hits in

---

31 http://www.aucklandcitylibraries.com/process.asp ?pageurl=/explore/teens.

32 http://web2.kclibrary.org/web2/tramp2.exe/goto/A0d2arns.000?&screen=PopularInterest.html.

33 Email from Quentin Chevignon, 19 June 2003.

chronological order is not a bad idea either, showing the latest information first works for most users.

6.3.5.2. *Conception and use of subject trees*

Apart from defining the conception of a subject tree, a task that remains unfinished, the principles of navigation within this subject tree are also rather unclear.

I used the term "Yahoo" earlier because it is used by suppliers. I believe that in English the expression "*browse through categories*" is used. In French, however, it is difficult to find an equivalent expression. The general public is generally unaware of the idea of an advanced search, i.e. a systematic search. Maybe a "topic-based search could suit the public[34]?"

"What about using the term portal? To me a portal means that there are three different possible ways of entering a search entry on the homepage. These are a search engine that is similar to *Google* and managed automatically, a subject tree similar to *Yahoo* that is managed by humans and specific links to favored sites. *Yahoo's* advantage is that the search engine function can be combined with the function subject tree at all levels of the subject tree. In other words, if I go down to the category 'medicine' I can use the search engine on Yahoo to only search this particular category. This type of combination might be very advantageous for us since a normal search engine like *Google* would create far too many hits[35]."

A keyword-based search [in MOCCAM] is an original idea as it is a full text search, i.e. the search is carried out on all words of the descriptions of records, which includes the summary and defines the relevance of a document according to how many times the keyword appears in the document. This enables the user to do exactly the same thing as most Internet users when using Google. They simply enter all keywords that they can think of. "The search engine then shows all descriptions of records that contain at least one of the keywords, which leads to too many hits. However, these hits are organized according to their relevance[36]."

6.3.6. *New solutions combined with traditional ones*

Finally, the main goal could be that users would be able to use their everyday language instead of the jargon of documentation. The latter therefore already needs to be spelled out in the form of a subject tree.

---

34 Email from Alain Caraco, 27 June 2002.
35 Email from Daniel Le Goff, 26 June 2002.
36 Email from Quentin Chevignon, 17 April 2003.

The hierarchically structured jargon, whether it is for classifications or a thesaurus, can therefore be used directly.

To apply this principle on a practical level, the easiest solution would be the use of traditional structured catalogs which contain search entries that are already spelled out for the user. "Of course this is only a short-term solution. The content management systems of libraries 'have to change their approach radically and develop a search engine that is similar to Yahoo'. This type of search engine needs to be entirely new and not only a direct link to traditional search tools[37]."

6.4. Conclusion

Apart from the changing the outer appearance of libraries, the real challenge is to introduce what librarians call bibliographic information into the information society that has been developing since the 1990s.

First of all, primary information, i.e. full texts, represents the matrix of change. Secondary information, i.e. information on information, metadata and catalogs, needs to be adapted to today's tools and applications.

There is no need to erase the past and start again from scratch. Tasks such as describing primary information, creating an index, classifying and finding information still remain the same. With the expansion of the Internet traditional documentation techniques become even more important.

The jargon used in documentation has also survived the adventure of new technologies. Online classifications such as the subject tree are based on traditional classification systems. However, other traditional search methods such as pre-established keywords are threatened by an entirely free search of all descriptions of records and, if possible, even summaries and contents tables.

However, we will try to save the keyword-based search. Let us try to do it with only two mouse clicks!

---

37 Email from Quentin Chevignon, 19 June 2003.

6.5. Bibliography

[BET 98] BÉTHERY A., *Abrégé de la Classification décimale de Dewey*, Cercle de la librairie, 1998.

[ERT 03] ERTZSCHEID O., *Chercher faux et trouver juste, Serendipité et recherche d'information*, CIFSIC, Bucharest, 2003. Can be accessed at thé following address: http://archivesic.ccsd.
cnrs.fr/documents/archives0/00/00/06/89/sic_00000689_01/sic_00000689.pdf.

[LAH 00] LAHARY D., *Vers une nouvelle bibliothéconomie ?*, Séminaire du Cercle d'études des bibliothécaires des régions Aquitaine-Languedoc, Nîmes, 6 May 2000, can be accessed at the following address: http://membres.lycos.fr/vacher/profess/conf/cebral-lahary.html.

[MUS 02] *Musique en bibliothèque*, Cercle de la librairie, 2002.

[USA 03] *Usagers et usagers des bibliothèques publique en Pays de France et Plaine de France*, Study report, September 2003.

[VOC 93] *Vocabulaire de la documentation*, 2nd edition Afnor, collection Les dossiers de la documentation, 1993.

Chapter 7

Stakes and Prospects of Heuristic Visualization for OPAC Use

7.1. Complexity of information systems

Libraries provide their visitors with aggregation systems that should simplify access to electronic resources. Libraries continually extend their collections by adding virtual sources. A website providing information improves the user guide in terms of its length as well as its quality [DUJ 06]. Electronic resources (bibliographic or full text databases) and online references (signets) are the new cornerstones of documentation. The "depth"[1] and the progressive change in these cornerstones enrich information systems and at the same time make them more complex, as all of this secondary information is needed for the overall organization of a library.

The way in which traditional resources such as printed documents are organized is also applied to digital collections whose restrictions, when it comes to access, are very different from traditional collections. Information systems are part of a new form of organization that unites printed documents, digital and digitized documents as well as descriptive data (Unimarc, MARC21, Dublin Core, etc.) in order to respond to the user needs as effectively as possible.

This increase in documentation requires attention and patience from the users when carrying out a search for information and documentation. The increase

---

1 The volume of electronic information within a database or a website is at first glance very difficult to define.

mentioned above is, however, taking place without considering the user [DEN 03] skills and knowledge of the global organization of libraries.

A search engine that does not meet the user needs when it comes to searching for information, even though it actively unites all visible or invisible elements that structure the library's space of multiplicity, represents a severe problem for the library as a whole. Necessary classification that structures the library's stock and allows for an extensive and associative understanding of it is now organized in only one directory. In other words, the physical space (the way people move through a library, reading rooms, etc.) and the virtual space (websites, portals, catalogs and electronic resources) are united in the same directory. The elements which are used to structure and organize this directory are largely underestimated.

The recent introduction of reading rooms into French university libraries which unite different related disciplines as well as how users move from one reading room to another might create a certain sensibility towards the proximity of different fields. This introduction does not comply with the rules that are usually applied to the strict organization of a library. This new set-up represents an improved access to collections. The architecture of the building also plays a significant role in the conceptual organization of a library that is based on a classification [PAT 04, RIB 96] system and the knowledge of librarians who are very familiar with the particular university they are working in.

Within the same institutions the topic-based redistribution of electronic resources according to the scientific courses offered at a particular university represents a supplementary organization. This topic based redistribution mirrors the classification system that is used and, of course, fully understood by librarians[2]. The advantage of this type of organization is the link between the needs of organization and classification and the queries users have when researching a particular topic.

Some of the reference points are metacognitive [ENG 01]. Remembering and understanding them helps the user to improve his/her search and at the same time develop a pragmatic, cumulative and operational knowledge of the library's organizational principles. As all information is united on one single screen it is much easier for the user to manipulate the complex and often very large amounts of information [LAC 01]:

---

2 At the University of Paris 8 electronic resources can be accessed by students and staff via the website of the university's documentation service. There is a topic-based access presenting different fields of science, which is based on the organization of reading rooms, as well as an access that is structured in alphabetical order (see http://www-bu.univ-paris8.fr).

… First of all we need to take the primary function of virtualization, the reduction of complexity, into consideration. The problem of complex digital data was resolved with the emergence of synthetic interfaces, i.e. virtual environments.

The aim is not to create an entirely virtual library as an ultimate solution [FEK 06], but to use this opportunity and offer the users the possibility to reconsider the library's services[3] in the current context of informational abundance where technical systems allow for access to all different types of information. As Michaël La Chance explains, this opportunity is becoming possible because complexity is being reduced.

7.1.1. *Complexity of inter-related information systems for documentation*

The developments which were initiated by IT and telecommunications have led to an increase in the complexity of information systems [PAP 06, NAR 05, DON 04, PED 01, LEM 99]. The never-ending possibilities of interconnecting technical devices (e.g. via web technologies), digitization and the standardization of bibliographic data have led to an increase in the level of complexity while offering access to sources of information.

The difficulties related to the overlapping of different fields of knowledge and research become even more complex with the inter-relation between different dimensions in documentation. Daniel Parrochia explained this issue in 1993 and until today no solution has been found to resolve it. "… decimal classification systems (Dewey, UDC) are arbitrary because of their undefined subdivisions. They do rather well when it comes to managing the exponential increase in the number of documents in a collection, but remain problematic as there are several possible interpretations of the inter-relations. Cognitive fields are infinitely overlapping each other and objects are fragmented and classified differently over the course of time. How can the overlapping of different fields of knowledge be overcome? Can it be overcome at all?"[4]

Despite the increasing number of inter-related information systems that indicate entirely new and very rewarding forms of reading[5], the real advantages of

---

3 "… Readers often request more search tools online, such as dictionaries, encyclopedias, grammar books etc. This type of usage and consultation might finally lead to a (re)discovery of real libraries…" Caroline Wiegandt, Scientific Council for Documentation, The Institute of Political Science of Paris, 8 October 2004.
4 Parrochia D., *Philosophie des réseaux*, Paris, PUF, 1993, p. 185.
5 For example, documentation system such as hypertexts: "Hypertext and other electronic information systems overcome human limitations by providing mechanisms for compact storage and rapid retrieval of enormous volumes of textual, numeric and visual data. The

documentation are hardly used[6] [CHA 99] as human beings have a limited capacity when it comes to the amount of information they can absorb [MIL 69].

7.1.2. *Complexity, training and catalogs*

For several years, and well before the beginning of the digital era, studies have shown that searching for information is not an easy procedure and that the methods, knowledge and skills needed cannot simply be acquired by a person trying to study these procedures on their own [DEN 03, PAG 02, BEA 02, ASS 02]. The results of these studies are underlined by the fact that the methodology of documentation is part of undergraduate courses as well as other forms of training offered to members of staff and other users of university libraries[7]. The recent phenomenon of digitization does not change this situation but highlights to what extent users are unable to master research (e.g. computer illiteracy) [VAL 05, FON 02, CHA 99]. Measures such as introducing training thus need to be taken in this field[8].

The large amount of documentation and the surplus of techno-documentation play a part in the increasing complexity. The interconnection of heterogenous information systems also plays an important role in this phenomenon of increasing complexity. This is why it is absolutely necessary to make cognitive elements that help users to find their way around and at the same time reduce the number of purely administrative applications that might confuse the user [FON 05, FON 02].

Our research project is dealing with the acquisition and the introduction of new material in university libraries. In this context, and specifically for the University of Artois and the University of Paris 8, for three years a new type of OPAC has been proposed to users[9]. It has been created to show that coherence and sense are the main pillars of a library. The library should be brought to the user's home [PAP 05]. The usage of libraries should therefore increase[10]. The installation of online catalogs

---

importance of the systems lies in their potential capacity to augment and amplify intellect" [MAR 88, p. 70].

6 "The void between the speed of knowledge acquisition and what is understood by the users who are able to access that knowledge on a technical and theoretical level is constantly becoming wider [HOL 99]."

7 The number of collections of articles on this topic created by the BBF is very high (http://bbf.enssib.fr). The websites CERISE and FORMIST offer help and feedback on anything linked to training in the field of information.

8 "… with the infinite capacity of storing knowledge using all different kinds of supports the role of our education system is to teach the art of how to navigate oneself and gain orientation in this field..." [BRE 95].

9 OPAC (*Online Public Access Catalogue*); this acronym is also used in France.

10 The feeling of resignation towards this wall of complexity needs to be prevented and the restless wandering of users through documentation centers must come to an end.

for some universities, i.e. the common documentation services for the University of Paris 8 and the University of Artois, are supported by interviews and questionnaires that enable us to gain an insight into the functionalities of online catalogs, as well as the user's technical needs. This manner of monitoring the usage of online catalogs also shows a form of "catachresis", rejection, diversion or simply support for this OPAC among users and librarians.

With the democratization of IT systems in libraries which has been taking place for the past 20 years, online catalogs have become an irreplaceable tool that enables a much larger number of users to access the library's collections [PET 03]. Online catalogs have become a mediation tool, an interface between collections and users of often very different backgrounds. Due to the technical progress of web technologies, online catalogs no longer work as alphanumeric terminals but have integrated applications of graphic screens, web navigators and hypertext. However, apart from technical[11] and theoretical[12] principles, hypertext is also based on associability created by graphic semiology [BER 67, BER 04]. Graphic semiology can be described as common denominators, social norms and semiotics when it comes to intellectual affiliation [COU 97], the increase of knowledge and an introduction of citizens to a specific society [SER 97].

7.2. Sense and visualization

7.2.1. *The multidimensional space of a library*

Numerous studies [POI 05, BPI 03, DUJ 85, LEM 90] show that the rules on how libraries are structured and function are unknown to their users. The negative image users have of a library is often based on an autodidactic approach towards libraries and documentation centers. Municipal libraries often neutralize reflexive approaches towards multiple layers of informational macro systems. They therefore do not create logical links between the different subject areas:

---

11 SUDOC for example condenses hypertextual mechanisms of links that are found in different technical manuals. Choosing one or several words that appear in the title leads to documents which are on the SUDOC index and whose title includes at least one of the keywords entered. The activation of a link representing a certain author also enables the user to obtain all works by this author contained in the SUDOC index. Other links describing the editor or the collection lead to documents that contain the name of the editor or the name of the collection.

12 See the works of Vannevar Bush (*As we may think*) and Douglas Engelbart (Augment/NLS) when it comes to the possibilities of increasing cognitive capacities by using hypertext.

– information systems[13]. Managing the structure of libraries is based on information systems of documentation (LIS) or integrated library systems (ILS). These IT tools supply data used for the management and acquisition of data as well as for the processing of documents [MAI 03]. They are integrated into the global IT systems of universities and digital working environments that are open to students, researchers and teachers;

– system of organized knowledge. Professional librarians working in the field of acquisition that is linked to scientific research[14] manage the collections and the global coherence of the documents in stock [JAC 04];

– the place where intellectual work is carried out and progress is made in personal knowledge and different ways of thinking. The diversity of the different ways in which research can be carried out as well as the multiple resources and documentation services make the library an instrumental system that combines intellectual, conceptual, technical and cultural aspects [FEO 99, COU 97];

– this social space [POL 01] of co-operation, exchange, meetings and discussions creates a link between the two main communities within a library, i.e. professional librarians and the users of a library.

The library unites different aspects (e.g. techniques, documents and cognitive aspects), approaches (e.g. organizational, educational, institutional and scientific) and features (e.g. social, cultural, economic and futurology). All of these aspects are integrated within the same global system of organization [DER 77].

7.2.2. *Accessing the stock of documents via metadata*

The digitization of information systems in libraries urgently requires standardization and guidelines in processing, organizing, describing and classifying the documents in stock. "Standardized descriptive data" is required to insure compatibility (e.g. by using the Z39.50 protocol). Data exchanged based on ILS can therefore be imposed rapidly. However, explanations on the functioning of these rules mainly respond to professional needs and do not make them more comprehensible for users who are unfamiliar with these technical and professional dynamics.

---

13 Online catalogs allow for partial visibility.

14 In the framework of common documentation services (CDS) representatives of different disciplines taught at the university, laboratories, groups and teams of researchers, and bibliographic requests by lecturers will all influence the institute's decisions in the field of acquisition.

Creating access to material while also explaining how the organization of a library works represents a challenge. This challenge requires the sorting and transmitting of essential knowledge from the abundance of information to enable the user to work with a search tool of his/her choice. Once users are able to use these search tools on their own they can choose a tool that corresponds best to their individual profile. This profile might vary for professional librarians, students, researchers, lecturers, etc. Their usage surpasses the level of "resourcefulness" that Martine Poulain mentions in the preface to Joëlle Le Marec's [LEM 90] book. This tool redirects the user in an ingenious as well as constructive way and therefore increases the properties that its creators have integrated into this tool.

In the field of ergonomics the notion of catachresis represents the fact that users are moving away from initial applications that were introduced by the creators of a tool and now use it in a different way. This change in usage is an indicator of the success of a project as the functions of a tool change [RAB 95]. The user might discover new functions that were not integrated into the tool intentionally. Different ways of carrying out research are given below. The list starts with very precise search entries and moves down to very vague ones:

– searching for a precise reference;

– searching for complementary elements in relation to a topic (authors, books, vocabulary, articles etc.);

– searching for a notion or concept that has still to be defined;

– searching for interdisciplinary proximity or topics that overlap one another;

– visibility of intellectual organization, i.e. an entire collection is being explored;

– using the catalog as a source of inspiration or for brainstorming;

– usage of the catalog to create a profile that will automatically update new information.

Even though this system is very complex, libraries always represent coherence [JAC 04] and the expertise of professional librarians can be guaranteed. This large variety of structures is the key to accessing collections. However, with the democratization of web access and the publishing of online catalogs, Unimarc structures, and classification systems such as LCC, Dewey and UDC as well as the main subject fields [MIN 05, LAH 97] are accessed by users who lack any kind of competence in this field.

Librarians therefore need to explain what has remained rather difficult or even impossible for the majority users to understand. Finding a way that enables all users to understand how a library is organized and which rules are applied could re-

establish a dialog with its public based on a better comprehension of how the library functions. Furthermore, the image of professional librarians could be improved by this pedagogical approach towards the library's users [CHA 05].

7.2.3. *Improved online catalogs – they lead to an increase in unintended applications*

The catalog plays a central role in a library and has more to it than the capacity to provide the required descriptive notes. It contains organizational information that reveals a distinct and empirical logic. Users often underestimate this logic as well as the importance of the actual catalog:

– architecture: the most recent library buildings make free access to collections easier. Even if the localization of a certain book is indicated in the descriptive note (precision might vary) it never explains to the user how collections are organized and why they are put into a particular room, or into different places;

– organization (and classification) as well as standardized descriptions of knowledge (RAMEAU, unified encyclopaedic directory on authorship and topic in alphabetical order). A classification system offers access that is based on different categories. This form of access is hardly used and difficult to introduce into reading rooms. Alphanumeric encoding is useful due to the fact that it is very concise, but, it does not provide any information about the variety and the intellectual organization of collections;

– policies on documentation reveal the library's system of acquisition and the value that is attributed to different collections.

Catalogs have, of course, benefited from the progress made in web applications. However, online catalogs still follow the traditional microfiche models. According to Christine Borgman this procedure no longer corresponds to user behavior when searching for information. " … we argue that most current online catalogs are based on card catalog design models, that this model does not map well onto online systems, and that the model is not based on information-seeking behavior" [BOR 96].

7.3. Visualization and the trail of knowledge

7.3.1. *Principles of a heuristic visualization*

"Similarly, imaging systems and computer graphics are further enabling man to see the unseeable and are becoming vital tools of research in some disciplines (...). Visualization is a method of computing. It transforms the symbolic into the

geometric, enabling researchers to observe their simulations and computations. Visualization offers a method for seeing the unseen" [VIS 91, p. 1].

Interactive visualization of information aims to provide a perception of information and manipulate this information. Its performance is based on the following aspects:

– the processing capacity of an algorithm is linked to efficient restoration of data on graphic screens[15] and the principles of creating two- or three-dimensional models of objects, interactions that include several different modes and, last but not least, virtual reality;

– the metamorphosis of the content represents a graphic representation of semantic and topic based links and inter-relations between the individual documents.

It is more precisely this axis of visualization that we are interested in as it allows for an elaboration of visual expression that is based on colors, forms, metaphors, lines, hierarchies, and composition [WOO 02]. This form of visualization calls for a convergence on the level of graphical semiology [BER 67], semiotics [STO 01] and visual communication that is needed to create a design for the given information [JAC 99].

"A less familiar but growing area is that of information visualization, which attempts to provide visual depictions of very large information spaces" [BAE 99, p. 260].

As a result, communication is based on infographics and interactive design that represents the respective information in order to emphasize its significant parts. The same effect cannot be reached within a written text[16]. We are talking about heuristic visualization and the four fields of knowledge mentioned above. Their aim is to add another level of comprehension to a text by using cognitive aspects of visual aids[17]. This especially applies to information that is already well structured and coherent but that simply cannot be understood on its own (see Figure 7.1).

---

15 Principles such as zoom focus on the context, hyperbolic subject trees, etc. [CAR 99].

16 The project *Understanding USA* is administrated by Richard Saul Wurman and is a very interesting example. It consists of a group of graphic and interactive features created by designers and architects of information. They were trying to present essential information from the annual USA report (1999) in a way that was easily understood by the country's citizens (http://www.understandingusa.com).

17 These images, no matter whether they are cartographic, monosemic or metaphorical, increase the potential for memorizing and learning [DEN 89].
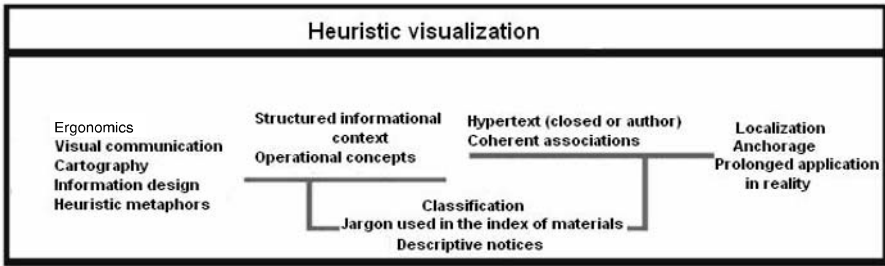
Figure 7.1. *Four components of heuristic visualization*

7.3.2. *Reticular systems and hypertextual trails*

The resources offered to the users of a library are covered by a diversity of different ways to carry out research. Obtaining the required information is directly linked to the special features of the individual's documentary research. Students, researchers or lecturers use this federated architecture and create their own trail of research and discoveries which fill the void between the subject of their research and the task of searching. These search tools provide documents and bibliographical information that should at least partly answer the user's questions. The fact that someone progresses in their research is just as important as the way in which his/her associations are structured. "... *understanding the coherence and the way in which hypertexts are organized is a fundamental step when trying to understand subject trees that represent a mental structure...*" [COL 04].

Systems of different types of logic are present within a library and can be represented as distinct reticular structures:

1) the reticular system is linked to the architecture of a library (Figure 7.2);
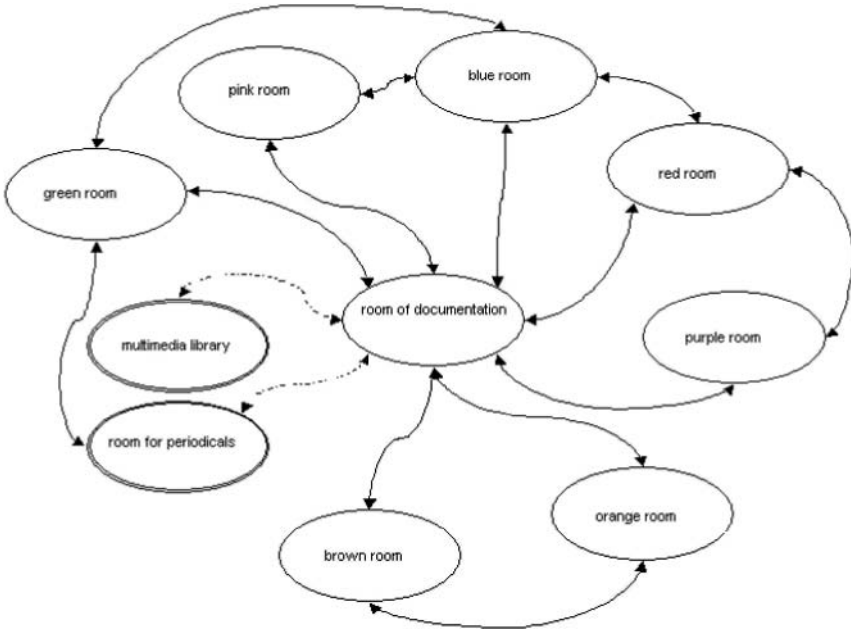
Figure 7.2. *Disposition, access and movement of users. Every reading room in the library of the University of Paris 8 corresponds to one or more disciplines. The way different disciplines are distributed is based on logic as well as on which rooms are located next to one another. As users move through reading rooms that are located next to each other the different disciplines start overlapping*

2) reticular classification systems. This system of organization is similar to subject trees. The user can first access the main categories, i.e. subject fields, and then go further down to specialized subdivisions. This system, however, also works on a transversal level and some books might appear in two or three different subcategories (Figure 7.1). This system allows for an approach based on different facets. "The basic principle of facet analysis is that concepts can be grouped using a characteristic of division which is not necessarily hierarchical. In other words, subjects which have previously been subdivided by progressive hierarchical arrangement, forming the familiar 'tree structures' of conventional indexing theory, can be looked on as patterns of horizontal division as well vertical divisions." [DUN 99, p. 133];

Title of a book	Classification mark	Label of subdivisions
Practices in communication from higher education to working environments	**302.22 MEY TEC**	Category of article: social interaction
	808.06 MEY	Specific techniques
Semiology of the image in advertising	**302.23 COR**	Ways of communication Mass media Social interaction
	659.13 COR	Particular categories of advertising
Large rivers	**551.48 BET**	Hydrology, lakes, rivers, streams
	628 BET	Sanitary and municipal techniques Urban engineering

Table 7.1. *Examples of transversal links between disciplines of some books (Dewey's system of classification, common services of documentation at the University of Artois)*

3) the RAMEAU reticular system. A description enables the user to move from one notice to the next and therefore change from one discipline to another following the principles associated with a thesaurus. The twofold semantic mark in bibliographic notes is not used often enough. It is based on the classification on the one hand, and on the headings and subdivisions on the other. Categories that are created within the classification system and concepts that stem from headings and subdivisions [MIN 05] enhance the quality of both simultaneously.

7.4. Interface, intermediaries and amplification of coherence

As the principle of visualization in this graphic interface (people are able to read and see) provides access to bibliographic data which is based on the presentation of elements that create coherence within a library, all results of a search (keyword-based or exploratory) are presented on the same screen.

The users can therefore exploit interactive functionalities that are part of the device and can set their preferences on one or the other reticular systems that is visualized as follows:

– synoptic graphics instantaneously deliver search results and a synthesis of different disciplines that are related to the search. The results are re-distributed into generic classification groups (Figure 7.3);
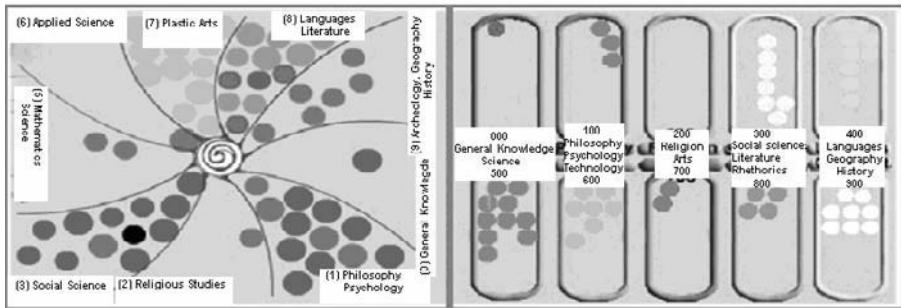
Figure 7.3. *Two graphic symbols that represent the distribution of results obtained in a search process throughout generic categories of UDC on the left-hand site (university library- University of Paris 8) and of Dewey on the right-hand site (university library – University of Artois)*

– localization synoptics that send the user back to the different categories, i.e. the different theme-based spaces in a library (Figure 7.4);
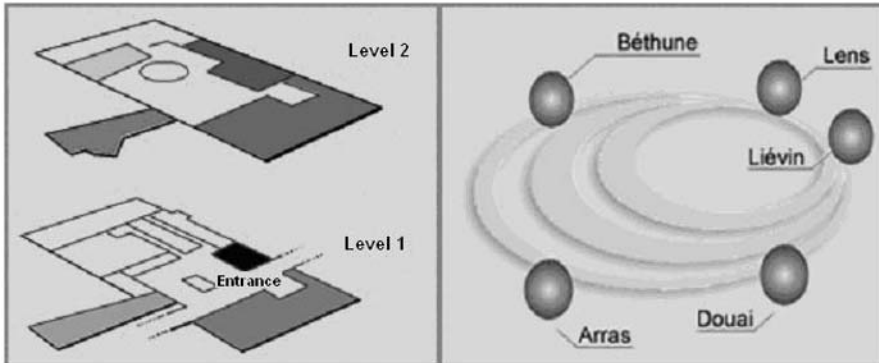
Figure 7.4. *Two cartographic representations. On the left hand-site there is the axonometry of reading rooms (university library – University of Paris 8), and on the right-hand site the five poles of the common documentation services in Artois (Lens, Béthune, Douai, Liévin, Arras) are shown*

– lists of titles, headings and subdivisions and classification put books in a semantic and cognitive context and make a difference between categorization (index of classification) and conceptualization (RAMEAU).

7.5. Usage and perspectives

The challenges of heuristic visualization that we have shown in the field of libraries largely surpass research that is currently carried out by students, researchers and lecturers. These challenges represent important changes in society as well as in the information society. The new trend towards an information society requires all individuals, i.e. citizens, to correctly use a variety of information systems and approach information in a responsible and critical way.

For several years scientific studies have been able to show the complex cognitive activity that is required in the research process. Today, the proliferation of information is based on the technological progress of information systems (directories, search engines, cartographic systems etc.) and certain online services can no longer be avoided by consumers (e-administration, call centers, banking, etc.). This trend adds new parameters to cognitive activity and makes the results that it should produce less reliable.

This complex cognitive level of researching therefore encounters advanced information systems (i.e. principles of usage and organization). The same applies to libraries that try to merge the search for information with documentation systems.

By using an "anthropocentric" [BEG 00] approach the focus was shifted to semiotic systems of a library which are vital for the comprehension and the approval of resources and the benefits they provide.

The devices that have been developed follow Edgar Morin's aim of "complex knowledge" which should "link, contextualize and globalize information and knowledge" [MOR 91, p. 456]. Studies that have been carried out at the University of Paris 8 since September 2004, as well as observations made in different units of documentation at the University of Artois, have shown an increase in research for information (bibliographic) which is due to the fact that several reticular systems have been united and can be accessed through one single interface. This study was based on questionnaires that were handed out to the library users (84 at the University of Paris 8 and 105 at the University of Artois) as well as data on the usage of these information systems (50,000 search entries at the University of Paris 8 and 18,000 for the University of Artois).

The headings and subdivisions in RAMEAU are hardly used and the classification[18] shows the confused and abstract knowledge users have when it comes to different entities in a search process. However, the intensive usage of hyperlinks between book → (classification, RAMEAU), classification → (book, RAMEAU) and RAMEAU → (classification, book)[19] is proof that users have a strong interest in categorization and conceptualization.

The possibilities heuristic visualization offers are immediate and significant as the amount of available information is continually increasing. Today's libraries have to adapt on a local as well as a global level [UNE 05, p. 67] and share primary data as well as metadata within an institutional network that is constantly growing. Libraries are offering visibility of their stock as online catalogs can be accessed via their website and enable users to consult and even order books without having to be physically present in the library[20]. Systems such as SUDOC (documentation system of French universities) make the bibliographic data of different university libraries available on their website[21]. These systems mainly focus on technology and do not respond to the needs of concepts and categorization that the users require. The fact that the users need concepts and categorization has been proven by the use of the features our tool offers. Due to these results two other universities are also going to

---

18 We have observed a strong imbalance between search entries on "headings and subsections in RAMEAU" and keywords entered into the fields "title" or "author". The possibility to explore the classification system is hardly ever used.

19 Hyperlinks are on average used for two types of search entries. The hyperlink RAMEAU → (title, classification) is the most frequently used one, directly followed by title → (RAMEAU, classification).

20 This applies to universities that offer the service of inter-library loans.

21 http://www.sudoc.abes.fr.

integrate this device into their portal. These are the University of Marne-la-Vallée and the University of Littoral. The introduction of this device enables every university to improve identification and access to categories of books.

7.6. Bibliography

[ASS 02] Assadi H., Beaudoin V., "Les annuaires du Web. Entre intermédiation neutre et choix éditorial marqué", *Réseau*, vol. 20, no. 116, p. 143-169, 2002.

[BAE 99] Baeza-Yates R., *Modern Information Retrieval*, Addison-Wesley, ACM Press, New York, 1999.

[BEG 00] Beguin P., Rabardel P., "Concevoir pour les activités instrumentées", *RIA*, *PEC'200*, 14, p. 35-54, 2000.

[BER 67] Bertin J., *La sémiologie graphique*, Gauthier-Villars, Paris, 1967.

[BER 04] Bertrand A.-M., "La bibliothèque et la ville', in 'Recherches et Bibliothèque", *Les Cahiers de Paris 8/Recherches*, p.117-120, PUV, January 2004.

[BOR 96] Borgman C.L., "Why are online catalogs still hard to use?", *Journal of the American Society for Information Science*, vol. 47, no. 7, p. 493-503, 1996.

[BPI 03] BPI, *"Les 25 ans de la Bpi. Encyclopédisme. Actualité. Libre accès"*, Actes du colloque international par la Bibliothèque publique d'information au Centre Pompidou, Paris, 23-24 October 2002.

[BRE 95] Breton P., *L'utopie de la communication*, La Découverte, Paris, 1995.

[CAR 99] Card S.K., Mackinlay J.D., Shneiderman B., *Information Visualization, Using Vision to Think*, Morgan Kaufmann Publishers, New York, 1999.

[CAT 01] Catellin S., "Sérendipité, abduction et recherche sur Internet", *Emergences et continuité dans les recherches en information et communication*, Actes du 12e Congrès national des SIC, UNESCO, 10-13 January 2001.

[CHA 99] Charon J.-M., *Réflexions et Propositions sur la déontologie de l'information*, Report to the French minister on Culture and Communication, July 1999.

[CHA 05] Chauvin S., Papy F., "Peut-on déranger le bibliothécaire à la banque d'accueil? ou comment rapprocher la communauté des usagers de celle des professionnels de la bibliothèque", *ACSI/CAIS 05*, University of Western Ontario, Canada, 2-4 June 2005.

[CHA 06] Chauvin S., "Les classifications instrumentées par l'hypertexte: améliorer l'exploration dans la structuration des saviors", *ISKO*, *Semaine de la Connaissance*, Nantes, 26-30 June 2006.

[COL 04] Colombi T., Baccino T., "Exploration visuelle et navigation dans les hypertextes: quelles stratégies?", *ERGO-IA 2004*, Biarritz, 2004.

[COU 97] COULON A., Le métier d'étudiant: approches ethnométhodologique et institutionnelle de l'entrée dans la vie universitaire, Doctoral thesis, University of Paris 8, 1990, 3 vol. The empirical part of this thesis has been published: Alain Coulon, *Le métier d'étudiant. L'entrée dans la vie universitaire*, PUF, Paris, 1997.

[DEN 89] DENIS M., *Image et cognition*, Presses Universitaires de France, Paris, 1989.

[DEN 03] DENECKER C., *Les compétences documentairs: des processus mentaux à l'utilisation de l'information*, Presses de l'ENSSIB, Paris, 2003.

[DER 77] DE ROSNAY J., *Le Macroscope*, Collections Point, Paris, 1977.

[DON 04] DONNADIEU G., KARSKY M., *La systémique, penser et agir dans la complexité*, Editions Liaisons, Paris, 2004.

[DUJ 85] DUJOL A., "Le clair et l'obscur: perception et usage de la classification par le public de la BPI", *Bibliothèque publique d'information, Centre Georges Pompidou*, Collection Etudes et recherches, Paris, 1985.

[DUJ 06] DUJOL A., "Les sites: trouver l'information ou la ronde des clics", *BBF*, no. 3, p. 38-42, 2006.

[DUN 99] DUNCAN E.B., "A faceted approach to hypertext?", in McAleese R. (ed.), *Hypertext, Theory and Practice*, p. 133-137, Intellect, Exeter, 1999.

[ENG 01] ENGEL P., "La cognition est-elle représentation", *Sciences Humaines*, hors-série no. 35, les Sciences de la Cognition, p. 73-78, 2001.

[FEK 06] FEKETE J.-D., LECOLINET E. (ED.), "Visualisation pour les bibliothèques numériques", *Document numérique*, vol. 9, no. 2, 2006.

[FEO 98] FEO A., "L'enseignement de méthodologie documentaire à l'Université Paris 8", *Documentaliste – Sciences de l'Information*, vol. 35, no. 3, p. 147-155, 1998.

[FON 02] FONDIN H., "L'activité documentaire: Représentation et signification", *Bulletin des Bibliothèques de France*, *BBF*, vol. 47, no. 4, 2002.

[FON 05] FONDIN H., "La formation à la recherche d'information: préoccupation citoyenne ou vision obsolète?", *Esquisse, no. 43,* July-August-September 2005.

[HOL 99] HOLZEM M., *Terminologie et documentation. Pour une meilleure circulation des savoirs*, ADBS Editions, Paris, 1999.

[HUS 98] HUSTACHE-GODINET H., Lire-Ecrire des hypertextes, PhD.Thesis, Université Stendhal Grenoble III, Presses Universitaires du Septentrion, October 1998.

[JAC 99] JACOBSON R., *Information Design*, MIT Press, Cambridge, 1999.

[JAC 04] JACOB C., "Travailler en bibliothèque", in "Recherches et Bibliothèque", *Les Cahiers de Paris 8/Recherches*, PUV, p. 31-39, 2005.

[LAC 01] LA CHANCE M., *Les penseurs de fer. Les sirènes de la cyberculture*, Editions Trait d'Union, Montreal, 2001.

[LAH 97] LAHARY D., "Que faire de RAMEAU ?: La bibliothèque en ligne", *Bull. Inf. Association des Bibliothécaires Français*, no. 174, p. 110-117, 1997.

[LEM 90] LE MAREC J., *Dialogue ou Labyrinthe: la consultation des catalogues informatisés par les usagers*, Etudes et Recherches, *BPI Centre George Pompidou*, 1990.

[LEM 99] LE MOIGNE J-L., *La Modélisation des systèmes complexes*, Dunod, Paris, 1999.

[LUP 03] LUPOVICI C., CLOAREC T., DE CHARENTENAY F., "Les usages de Gallica", *Bulletin des Bibliothèques de France*, vol. 48, no. 4, p. 40-44, 2003.

[MAI 03] MAISONNEUVE M., *Du catalogue de la bibliothèque aux ressources du Web*, ADBS, Paris, 2003.

[MAR 88] MARCHIONINI F., SHNEIDERMANN B., "Findings facts vs. browsing knowledge in hypertext systems", *Computer*, January 1988.

[MIL 69] MILLER G.A., *Psychology. The Science of Mental life*, Penguin Books, Harmondsworth, 1969.

[MIN 05] MINGAM M., "Rameau: Bilan, perspectives", *BBF*, no. 5, p. 38-47, 2005.

[MOR 91] MORIN E., "Introduction à la pensée complexe", *ESF*, 1991.

[NAR 05] NARBONNE Y., *Complexité et systémique*, Hermes, Paris, 2005.

[PAG 02] PAGANELLI C. (ED.), *IHM et recherche d'information*s, Hermes, Paris, 2002.

[PAP 05] PAPY F., CHAUVIN S., "Pour une approche visuelle et ergonomique dans la recherche et l'exploration d'informations au sein d'un OPAC de SCD", in Papy F. (ed.), *Les bibliothèques numériques*, Hermes, Paris, 2005.

[PAP 06] PAPY F., "Modélisation des metadonnées bibliographiques: une contribution insrumentée destinée à améliorer la circulation des usagers dans les connaissances de la bibliothèque", *IC 200*6, Biarritz, 2006.

[PAT 04] PATY A., *Une bibliothèque: la bibliothèque de l'Université à Saint-Denis, de Pierre Riboulet, Architecte*, CAUE 93, Bobigny, 2004.

[PED 01] PEDLER E., ZERBIB O., "Les nouvelles technologies à l'épreuve des bibliothèques", *Etudes et Recherches, BPI*, November 2001.

[PET 03] PETERSON BISHOP A., VAN HOUSE N.A., BUTTENFIELD B.P., *Digital Library Use: social practice in design and evaluation*, p. 201-205, The MIT Press, Cambridge, Massachussetts, 2003.

[POI 05] POISSENOT C., "Usages des bibliothèques: approche sociologique et méthodologie d'enquête", *Les Cahiers de l'ENSSIB*, 2005.

[POL 01] POLITY Y., "Les bibliothèques, objets de recherche universitaire", *BBF*, vol. 46, no. 4, p. 64-70, 2001.

[RAB 95] RABARDEL P., *Les hommes et les technologies: approche cognitive des instruments contemporains*, Armand Colin, Paris, 1995.

[RIB 96] RIBOULET P., "Le caractère du bâtiment", *BBF*, vol. 41, no. 5, p. 72-79, 1996.

[RIC 01] RICE R.E., MC CREADIE M., CHANG S.-J.L., *Accessing and Browsing: Information and Communication*, The MIT Press, London, 2001.

[SER 97] SERUSCLAT F., De l'élève au citoyen. Rapport sur les nouvelles techniques d'information et de communication, Office parlementaire d'évaluation des choix scientifiques et technologiques, rapport no. 45 (Assemblée nationale), no. 383 (Sénat), Paris, September 1997.

[STO 01] STOCKINGER P., *Traitement et contrôle de l'information: procédures sémiotiques et textuelles*, Hermès, Paris, 2001.

[UNE 05] UNESCO, *Vers les societes du savoir*, Rapport Mondial de l'UNESCO, Editions UNESCO, 2005.

[VAL 05] VALADEZ OLGUN R.G., PAEZ ARANCIBIA J., ZAPATA GUERRERO M., "Management in the use of digital infomartion resources: case study at the library of the biblioteca de la Facultad de Estudios Superiores Cuautitlan UNAM", *ISKO 05*, p.613-624, Barcelona, 2005.

[VIS 91] VISVALINGAM M., "Visualisation, ViSC and Scientific Insights", *CISRG Discussion Papers Series*, no. 9, University of Hull, 1991.

[WOO 02] WOOLMAN M., *Données à voir, le graphisme d'information sur support numérique*, Thames & Hudson, Paris, 2002.

This page intentionally left blank

Chapter 8

3D Interaction for Digital Libraries

8.1. Introduction

In the chapter "Inventions à Faire" in his 1934 book *Traité de Documentation* Paul Otlet describes a process which prefigured what ultimately became the World Wide Web [OTL 34]. The similarities between his work at that time and current research programs used in digital libraries are staggering:

8 – Tele-reading – an application similar to the television. First of all it enables the general public to read texts from far and wide. Secondly, with the help of an appropriate device it enables people to read these texts which are made available for this particular reason. Thirdly, it allows people to have access to texts from books which are available on the shelves of a library or from pages grouped together on folders. (Service: to increase the availability of library collections and documentation.)

It has not only been television which has been able to use this founding idea by Otlet, but also the Web and the Internet, both having experienced an enormous level of success. Nevertheless, if the Web and its associated technologies (formats of metadata description, exchange systems and peer to peer systems) enable us to read hundreds of thousands of digitized works which are currently available online, it must be pointed out that the thinking behind such a creation was not a particularly easy task.

Chapter written by Pierre CUBAUD.

The display of non-quantitative information is a recent research development in the field of informatics which has been improved by the progress made in graphic display hardware and which has been steadily improving due to the rapid growth of information which is available in digital format. Numerous devices and various pieces of human-machine interaction software have been developed in order to help the user understand the large number of texts organized together within a network (such as hypertext and research results). However, there still remains much to be done in order to evaluate the potential of such software and devices in the context of digital libraries [CHE 02]. As the title of the report suggests, we will only deal with the study of the potential of 3D interaction techniques in texts which are based on the founding text by Robertson *et al*. [ROB 93]. Other processes exist and we will come back to these ideas when dealing with collections of articles [CAR 99] and [CHE 99].

In this chapter we describe the experiments carried out in this field at the National Conservatory of Arts and Crafts (Conservatoire national des arts et métiers[1] or CNAM) over the last decade. These different experiments have enabled us to progressively specify the basic components of a 3D workshop for the reading of ancient collections which have been digitized. We will also discuss how they function. The technical aspects linked to the programming of the interaction in a 3D context have been excluded at our discretion: a talk on this can be found in [CUB 01] and especially in [TOP 02]. We have tried to place this piece of work together with other similar pieces of work that have been carried out elsewhere. However, this work cannot be deemed to be exhaustive. This report is organized into three parts. First of all, we describe the techniques which are used and that enable us to recreate the 3D characteristics of a book. We then discuss the current limits of the Web as an interface for people to have access to digitized books for two reasons: reading the books and then carrying out research within the libraries' collections.

8.2. The page as a surface

The digitization of books is carried out on an almost industrial scale in powerful machines after the books have been guillotined and placed flat into the automatic chargers. Without the guillotine process the books would be distorted because of their shape. In fact, the surface to be digitized is rarely flat because of the presence of the spine of the book. This is why the books are generally placed under a pane of glass which forces then to become flat. However, it is believed that this technique would not be appropriate for precious works and that it would be out of the question to manipulate these books in such a way. Furthermore, the need to flatten the book

---

1 In collaboration with G. Deblock, J. Dupire, D. Girard, A. Topol and P. Stokowski.

against a pane of glass considerably slows down the process and makes it difficult to move the book. Other machines are therefore used which digitize the books as they are, without using a pane of glass. Recently, certain machines have been equipped with an automatic page turner (4DigitalBooks). The distortions of the images are processed by the software by following different problem solving techniques, such as the rectification of the edges of the book [BRO 03], the lines of text (the software I2S Book Restorer), inclinometry [WAD 95], etc.

As soon as the book itself becomes the focal point, the approach mentioned previously becomes insufficient. Examples would be artists' books with numerous superimposed collages, as well as animation books with moving parts and even folders made up of specimen sheets of herbariums. As a result, it seems only natural to try and capture this geometry by 3D digitization. This could also lead to openings with regards to rare but conventional books in e-commerce, a field which is currently experiencing a growth period. The process of 3D digitization has also led to the creation of an application which can be used in the digitization of documents which have extensive damage. The example in Figure 8.1 shows how we reproduced a volume which was damaged by humidity, taken from an ancient collection at the CNAM.

Figure 8.1. *An example of a volume from a collection damaged by humidity: Privat, Deschanel, Précis de physique, Paris, 1855, in-8°*

Despite the increasing systematic use of 3D digitization in the area of culture and the development of a specific market, at present there is no hardware or material that exists which is dedicated to the 3D digitization of documents. Two techniques of digitization have been studied in this chapter.

8.2.1. *Structured light*

A luminous design, for example a grid, is projected onto the surface and a shot is taken from a different angle. The resulting distortion of the design on the image enables us to physically reduce the size of the surface. The resulting surface size is linked to the size of the design. The different lines of the grid have to be individually identified, as there can be ambiguity from time to time. In order to distinguish between the lines of the grid, each line can be given a different color. This is the principle used in the Minolta 3D1500 device which was used in the Digital Atheneum Projet [BRO 00] for very wrinkled manuscripts dating from the Middle Ages. The camera was also used in a test at the CNAM in June 2000. The findings which resulted from this test in June 2000 were applied to a binding process known as 12-inch binding which can be seen in Figure 8.2. 12-inch binding was a common process used in the binding of books throughout the 18th century in France (Nollet, *Leçons de physique*). The wrinkles of the paper (rag paper in the 18th century and relatively thick), as well as the spine of the bind, have been re-created very well. However, it is also apparent that the rough edges of artifacts and the crinkles of the edges have been smoothened due to the interference between the strips and the bundles of pages. The 3D1500 digital camera is no longer available on the market.
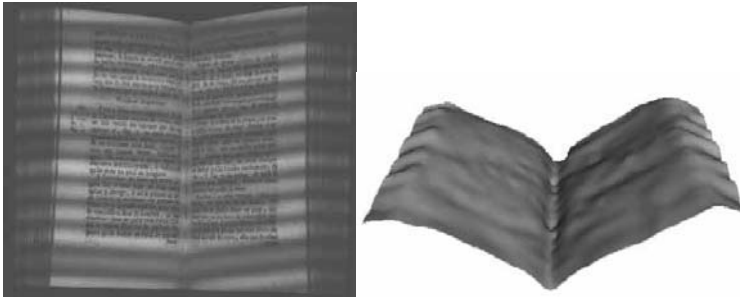
Figure 8.2. *Results from the tests carried out using the Minolta 3D1500 digital camera: on the left you can see the projection of strips of color and on the right you can see the resulting surface*
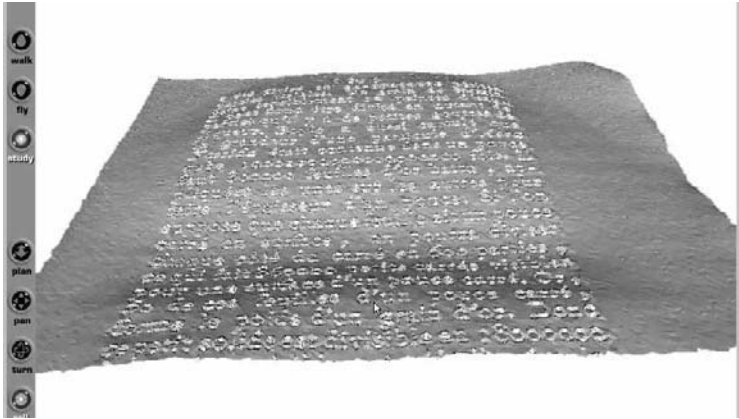
Figure 8.3. *The surface obtained with the VIVID 700 (20,000 triangles)*

All-in-one devices known as 3D scanners also exist. They are made up of a scanning laser linked to a digital camera. In this case, it is more a question of using structured light where the projected line is generated by the scanning light (in a view perpendicular to that in Figure 8.2). The projected line scans the object (in the view that can be seen in Figure 8.2). We have been able to test the Minolta VIVID 700 scanner (which functions according to this principle) on an ancient book similar to the one mentioned earlier (Lolande, *Almanach de physique*, 12-inch binding). The texture of the printed characters is interpreted as being smooth (Figure 8.3): in fact the device only works with a radiometry of the uniform object. However, the accuracy of restoration is extremely good in the other zones: smaller than 1/10 mm. A more economical device is described in [BRO 01] in which the documents are scanned by a luminous line and the images are photographed one after the other. The resulting accuracy is similar to that achieved using a laser.

8.2.2. *Photogrammetry*

Photogrammetry allows for the reconstruction of the surface from a couple of stereoscopic photographs [EGE 01]. Each couple of points identified by the correlation between the two images gives the coordinates of the corresponding 3D object point. In relation to the techniques mentioned earlier, photogrammetry has several advantages such as: coupling between the acquisition of the texture and the acquisition of the surface, the total separation between the acquisition phase (rapid) and the processing phase (slow), the robustness of installation. General photographic hardware which is approximately positioned and which is non-standardized is

already leading to positive results. If we continue with the standardization of the equipment it is possible to obtain definitive metric information directly onto the surface of the document (it also becomes possible to use a side view, for example). On the other hand, photogrammetry has two main disadvantages: the need for the presence of a texture on the document, for uniform areas with no varying textures the process of photogrammetry cannot be used (margins, spaces between lines, etc.), and the risk of false correlation, which remains the major problem of photogrammetry if we hope to use it on its own without the interference of any human control.

The view of the document has some characteristics which distinguish it from the most common applications (for example, aerial or architectural photogrammetry). Certain aspects make the process of 3D reconstruction easier: the external orientation of the shots is fixed and known; the object is normally relatively flat without any discontinuities or occlusions which facilitates the correlation between the images. However, the focus is close-up and variable which makes any preliminary standardization quite difficult. The strong convergence leads to problems in terms of the depth of field (DOF). Books which possess a texture with a large repetition of images (for example, the letters in a text or the features of engravings in illustrations) also carry the risk of creating false correlations. In [CUB 04] we describe the organization of a photogrammetric chain which takes these specifications into account.

This chain was tested on several samples by using two cameras which are available on the market to the general public. For 4Mpix images the resolution of the device is expected to be 0.1 mm. For the test book which is shown in Figure 8.1, after approximately five minutes of the photogrammetry process the resulting surface of the document is composed of 30,000 triangles (Pentium 4.2Ghz, Java application), which can be seen in Figure 8.4. False correlations represent 0.6% of the total triangles. The false correlations are, however, very awkward because they lead to the presence of unacceptable surfaces on the document (Figure 8.5). Several problem solving techniques are currently being investigated in order to reduce this rate. Other techniques of surface meshing are also being studied which are better adapted to non-text documents (engravings in particular). In addition, it should be possible to reduce the processing power through parallel processing in order to use such techniques in industrial applications.
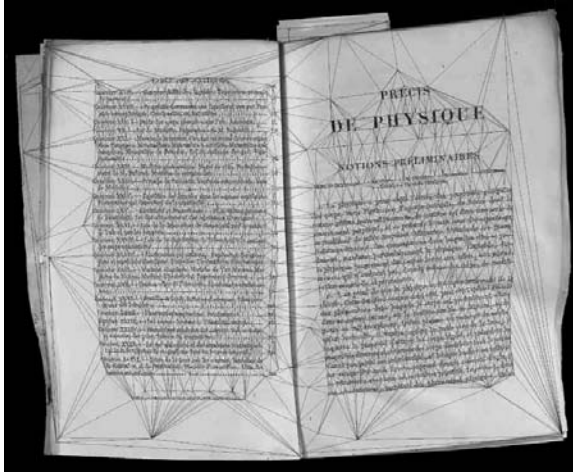
Figure 8.4. *Left-side view after triangularization (27,075 triangles)*

From the information that we have about the 3D surface of the digitized document we can try and unfold the document or make it flat. This is useful for bound manuscripts such as parish registers or the manuscripts of important authors. The problem solving techniques used to rectify the lines of texts cannot be used here. Historically, this is the first application considered for 3D digitization [BRO 00, PIL 01]. The progress which has been made in terms of 3D visualization in real time enables us to offer the user a strictly 3D consultation interface. The use of a general 3D navigation tool (in VRML, Metastream or QTVR format) is rather tedious and boring in terms of the manipulation of documents. In fact, only the zoom actions and the rotation around the axis defined by the width of the book's binding are very useful. For the consultation of an entire text we also want to be able to offer the possibility of scrolling through pages. The 3D object can then be placed within the more general 3D interface of the digitized library. We will investigate these issues in the forthcoming sections of this chapter.
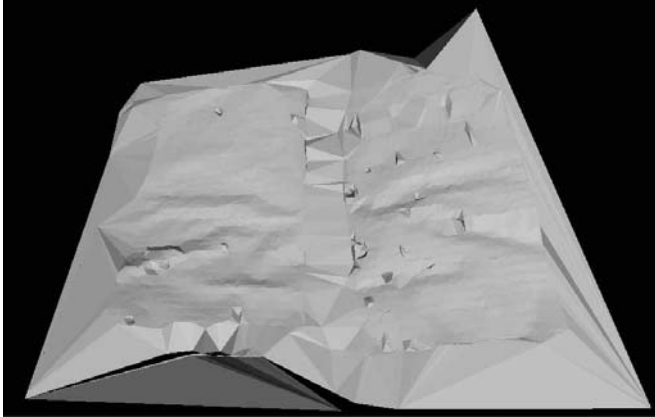
Figure 8.5. *3D surface translated into VRML, unflattened text in order to remove the false correlations*

8.3. The book and reading interfaces

The interfaces of digital libraries currently rely on HTML and associated technologies. We will take, as examples, the interfaces that we helped create and set up within the CNAM. These are obviously neither the most important, nor the most sophisticated available, but they are representative to us. Launched in October 1993 on the first French website, the Association of Universal Booklovers (ABU, http://abu.cnam.fr) has classic works taken from the French speaking world which are available online. The books are digitized in "text mode" with minimal tagging, similar to that of Project Gutenberg in the USA [CUB 98a]. The Digital Conservatory of Arts and Crafts (Conservatoire numérique des arts et des métiers or CNUM) (http://cnum.cnam.fr) has been online since January 2000 [CUB 03a, ROZ 0]. The result of a partnership with the central library of the CNAM and the Center for the History of Technology (Centre d'histoire des techniques (CDHT)), the website today has 300 reproductions of ancient scientific and technical books in facsimile format.

This type of literature has several interesting characteristics from an editorial point of view. The illustrations in them play an extremely important role. They are generally grouped in separate bookshelves (up until the middle of the 19th century) and these bookshelves have a large variety of formats and sometimes a large variety of formats within the same volume. The editorial structure is generally very sophisticated (as opposed to the novel, for example), but is also sometimes incoherent when the publication is spread out over a large number of years. The

quotes and notes also play a key role in academic-type publications. Sequential reading is therefore relatively rare, and these books are rarely read on their own.

We have tried to take this information into account in the definition of the interface of the CNUM:

– the user can easily open several windows in order to navigate amongst several facsimiles of texts or search through their tables;

– zoom functions enable the user to modify the current page to the size of the window which contains the page;

– specific navigation tools have been created in order to move from shelf to shelf or to go directly from any page to a particular shelf mentioned in this page.

These functions have been created by scripts on the server and by dynamic HTML pages without having to resort to an exterior plug-in (which is a fixed constraint in our specifications). Figure 8.6 is a screenshot of a hypothetical work session dealing with Pascal's Arithmetical Triangle. We can find a page from the CNUM catalog (a), as well as two windows (b) of which one has been reduced by the user. Another window corresponds to another of Pascal's texts on the site of the ABU (c). The user also takes notes or extracts of text with his preferred software (d).

Figure 8.6. *Print screen of a working session in progress with the CNUM and ABU websites*

Despite these efforts it seems that the majority of users of digital libraries download their documents in order to print them. Letters from users of the CNUM have mentioned that the main problem they have is related to the downloading and

printing of documents. In [CUB 98a] there is a comparative study for the site of the ABU which compares the downloading of documents to the reading of documents on their website. The report shows that on average the size of one download is equivalent to eight documents being read. The study "BibUsages" carried out on the Gallicia website has shown a more unfavorable result for the scanning through and reading of documents: for every 548 documents that were downloaded only 314 were read. Web interfaces have only experienced a limited amount of progress in comparison to FTP interfaces which were used previously.

Interfaces other than the Web are currently being used for the consultation of facsimiles of collections. The most widespread is Adobe Acrobat, but its use for large digitized collections with images remains difficult to master (navigating within illustrations is not very useful in this context, while researching occurrences of illustrations is also slow). Experimental systems have been developed in several laboratories: for example, BAMBI [CAL 98], Philectre [LEC 98] and DEBORA [DEB 02]. Like the Internet, all of these interfaces function based on the basic elements of the WIMP paradigm (Window, Icon, Menu, Pointer). The limits of this process arise in terms of the amount of space required on the computer screen, as can be seen in Figure 8.6. It would be difficult to add another window for supplementary text. Furthermore, even the organization of the window in itself involves a lot of work in terms of clicking, moving the windows and icons.

The effort forced on the user of such a system by reading on screen has been analyzed in [OHA 97] in comparison to traditional reading. This work has highlighted the need to have more fluid navigation techniques for reading interfaces as well as a larger flexibility in controlling the organization of documents on screen. It seems that techniques such as those commonly used in computer-aided design can help resolve the issue. In bringing together the 3D geometric and visual characteristics of the documents and by inserting them into a common 3D scene, it then becomes possible to manipulate them.

In the reconstructed interface in Figure 8.7 the 3D scene is limited vertically by the floor. The camera is permanently fixed above this floor and its field is also fixed. The facsimiles are presented in a specific tool in the shape of a tripod (similar to a simplified lectern). The book can be positioned arbitrarily on the floor, pushed back, pulled or pivoted by the user who can give their commands by selecting them on the green bar at the bottom of the tripod. This manipulation is carried out in real time with a simple pointing device similar to a mouse: two degrees of freedom are enough to determine the position of the tripod on the ground. Several tripods can also be positioned on the work screen. The movement of a tripod can be restricted due to the presence of other tripods or if there are any collisions between these 3D

objects. The tripod can be removed by clicking on the green bar with a right-click of the mouse. It is possible to develop other interactive shortcuts to enable the user to move the tripod to the foreground of their computer screen. These shortcuts can also be found at the base of the tripod as buttons or menus.

Figure 8.7. *The reading tripod*

Scrolling through pages of the books is made possible by clicking on the page you are currently reading on the screen. Just like a traditional 2D window the tripod has a lift on its vertical axis in order to speed up the progression through the document (this is represented in Figure 8.7 as a ball). The horizontal axis enables the user to control the connection between the dimension (in pixels) of the facsimile and the chosen dimension of the tripod. There is also a zoom function which is independent of the positioning of the tripod that the user is working with. The size of the tripod can also be adjusted by using a specific handle (represented in Figure 8.7 by a ball at the top left of the tripod).

The thickness of the book should also be taken into account. The positioning of the book allows for easier localization of the selected page in the global text. The sections can be added as bookmarks corresponding to chapters or to passages selected by the user as in the 3Book system [CAR 04]. It is also possible to modify the appearance of the tripod in order to be able to display the open book. The action of page turning can be simulated by moving from one edge of the selected page to the other (Figure 8.8). A similar process (non-3D) had been suggested in the experiment known as "Turning the Page" carried out by the British Library [CAR 98]. This experiment is regularly used with great success in general public

exhibitions of precious texts (for example, the exhibition of Leonardo da Vinci's manuscripts at the Louvre). A more sophisticated animation based on the physical characteristics of a bent piece of paper has been described in [CHU 04, WIT 03]. The page-turning associated with a 3D representation of pages remains to be studied.
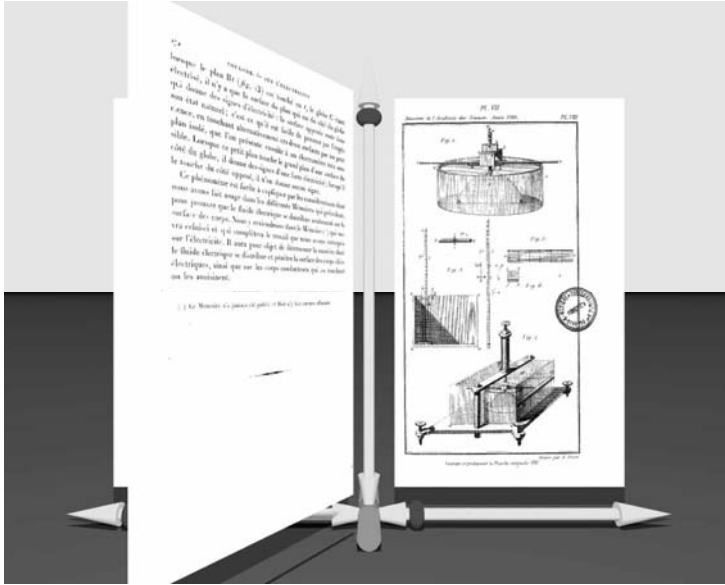
Figure 8.8. *The tripod in page-turning mode*

The proposed design for the lectern is clearly arbitrary. Due to the performance of the 3D structure we have limited the design to basic geometrics, and the analogy with a conventional window on a computer screen. The main advantage of the device (in comparison to those suggested in [CAR 04] and [CHU 04]) makes it possible to have a large number of lecterns within the same work session. The device also completes the lecterns by adding research and navigation tools to the collections of books.

8.4. Research collections and research interfaces

Figure 8.9 shows the interface of a piece of research carried out in three well-known libraries: the CNUM (CNAM), Gallica (BNF) and the Memory of America

(Library of Congress). None of these could satisfy a contemporary reader like Alberto Manguel [MAN 98]:

> Whenever I choose a book to read in bed or to place on a lectern, to give to someone or to read on the train, my hands take into consideration the shape as much as the contents of the book. Depending on the occasion or the location where I have decided to read I prefer something small and intimate, or ample and substantial. Books assert themselves because of their titles, their authors, their positioning in catalogs and libraries, the illustrations on their covers and also by their size. Depending on the time period and place, I have become accustomed to seeing books appearing in all shapes and sizes, and as in all types of fashion these ever-changing aspects add a certain characteristic to the definition of what a book is. I judge a book by its cover, I judge a book by its shape.
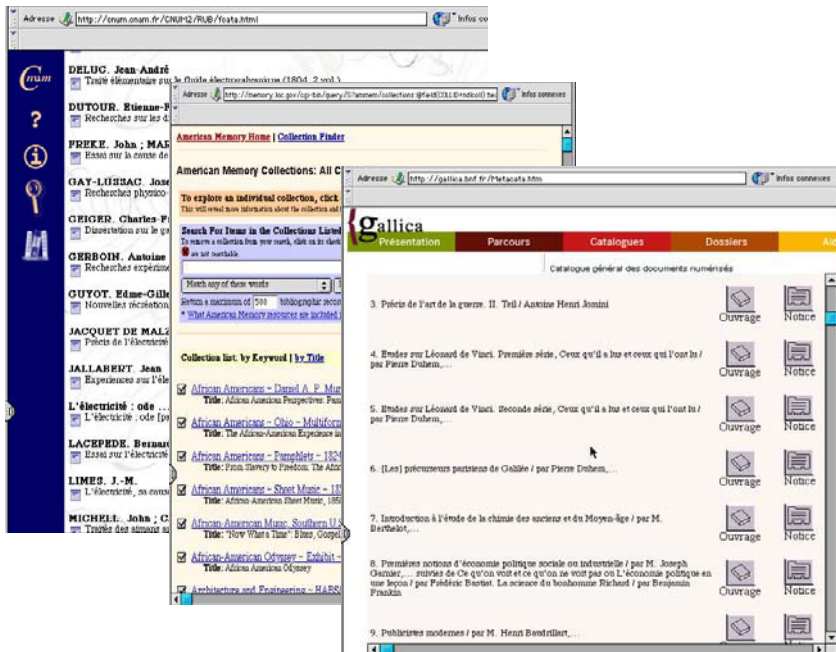
Figure 8.9. *An extract from three digitized libraries (pages of catalogs)*

In practice, the search mode is always associated with a search engine, a function which has been judged as essential as soon as the corpus reaches a certain level. As sophisticated as this function is, it is only useful for the user who already has a preconceived idea of what they are going to research: it is therefore not possible to lounge around in a digital library.

It is quite easy to complete the text interfaces of research tools by adding a much more graphic function, for example by digitizing the spine of a book to its respective dimensions, as can be seen in Figure 8.10. However, for the websites mentioned earlier this information was not gathered during the design of the websites. It is also believed that in most cases it is not possible to distinguish between books just by looking at the cover. From information that relates to the physical make-up of the books (page numbers, publication dates) it would perhaps be possible to construct a more abstract interface relying on a "grammar" of predefined shapes for which an inventory would need to be established. A similar approach was used in the Libviewer project [RAU 00]. In this system, the metadata is accessible and the result of the research requests is displayed in the shape of 3D bookshelves. These interfaces have the disadvantage of not being able to simultaneously display a large number of volumes. Of course, a zoom function is feasible but its use in a 2D graphic context does not allow for fluid navigation. Here we meet the limits of the WIMP process mentioned in the previous section.

Figure 8.10. *An alternative visual interface (http://jasmin.cnam.fr:8081/REL/)*

Carrying out a search in a library using virtual reality techniques has been investigated by several teams. One of the first 3D representations of an existing library (with VRMC) was created at The Institute of Research and Musical/Acoustic Co-ordination (L'Institut de Recherche et Coordination Acoustique/Musique or IRCAM) [FIN 00]. Similar experiments have taken place at the National School of Mining Engineering in Nantes [PLE 01] and at the University of Karlsruhe [CHR 02]. In these interfaces the organization of the actual library is respected (study and work rooms, shelving areas, even corridors and stairs). Beyond their undeniable asset as a communication tool, such processes also have limits. Navigation from one

virtual room to another is time-consuming (it is impossible to imagine how such an interface would represent the François Mitterrand library in Paris). Furthermore, all of the collections which can be digitized are not strictly accessible to the public, nor are they linked to a specific establishment. In [FOX 97] a summary is given of the experiments carried out for the 3D navigation on a national scale of a large corpus of digitized theses.

In [CUB 98b] we have presented an interface of the same type for users of the CNUM. The users of this library can navigate in VRML scenes which represent large virtual bookshelves. Several structures have been studied: mainly linear or helicoid structures. A cylindrical-shaped structure is all that is required to ease navigation within a scene. In such a scene (Figure 8.11), the user can pivot the view point and move it in order to zoom towards the bookshelves. The organization of the collection of books can also be reconstructed according to the user's needs. In this case the view point is positioned to the books or collections of books corresponding to the user's request, whilst the remainder is placed "behind" the user. For requests with two criteria (for example, date of publication = 18th century AND subject = electricity) rotating the view point shows the books which answer the first criterion whilst the vertical axis enables the user to navigate through the collection according to the second criterion. These two movements can be easily carried out with an ordinary mouse.
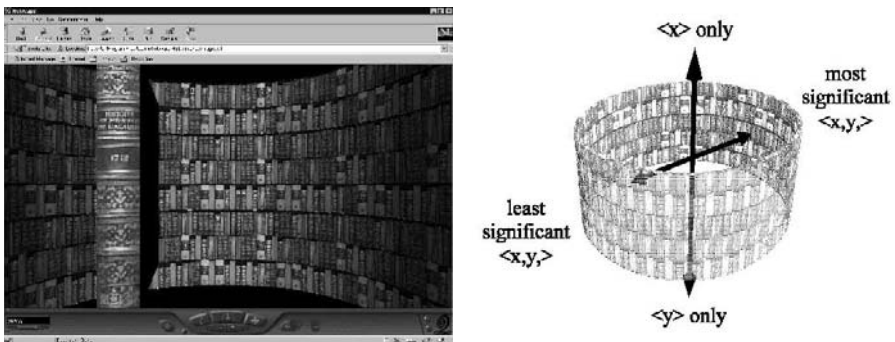
Figure 8.11. *A VRML scene for virtual shelves. The selected book appears in the foreground (on the left). Criteria of shelf classification (on the right)*

The virtual bookshelf can be inserted into the general reading interface in several ways. However, a compromise must be found in order to try and avoid too much information appearing on the screen at once (which tends to disperse the objects from the scene), and reduce the navigation time in the scene (which gets longer if

the objects on the shelves are far from one another). Figure 8.12 is a reproduction of the organization of the 3D scene which is quite similar to that of Web Forager [CAR 96].

Figure 8.12. *The entire collection is shown and accessible at the back of the screen. The favorite books or basket are available in the foreground*

The collection of books is always visible in the background and zooming into the bookshelves does not affect the positioning of the lecterns. It is understood that with this layout the screen overloads as soon as some of the lecterns are activated and a large number of the bookshelves become invisible. A wider scene therefore needs to be defined. In Figure 8.13 the zone of movement of the lectern is organized in the shape of a disk around the user. By rotating the lectern 180° on the Y axis the screen moves from the area of the lectern to the area of the bookshelves. There is also the possibility of positioning the lecterns in front of the bookshelves. Other organizations are possible. In [CUB 01] and [CUB 02] alternative designs have been suggested.

A demonstration (based on Opengl) was presented to NITC professionals during the JFT 2003 and ECDL '03 conferences; see [CUB 03] and [DUP 03] respectively. A wider audience was attracted to the general public exhibition known as Image by Image (Montreuil, March 2003) and to the National Science days in which the CNAM participates. At each of these exhibitions we had a stall dedicated to the 3D workshop and another stall which enabled the general public to have access to the

CNUM website. The sessions were limited to a few minutes by a timer and the users' actions were recorded. From these demonstrations we have been able to conclude that the users are able to use and manipulate the lecterns within a short period of time. The page-numbering function was widely accepted by the general public and by librarians (but not always by professional computer scientists). Some modifications from the designer have been necessary. The extent to which the lecterns could be moved was limited: a lectern cannot be larger than the screen nor can the lectern be minimized on the screen. The detection of any collision between objects has improved, as has the lighting and the reproduction of shadows.
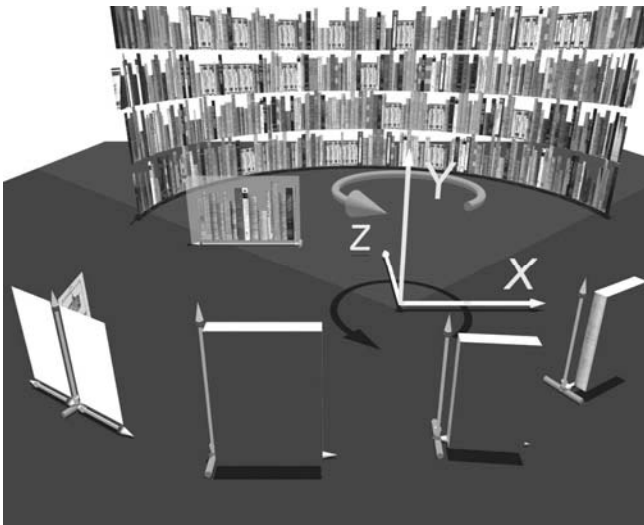
Figure 8.13. *The organization of a workspace for reading and for carrying out research within a collection of books*

8.5. Conclusion

Several functions have still to be implemented from our workshop which aims at reproducing documents in 3D. In fact, the current demonstration only includes one simplified collection, the creation and removal of lecterns and page turning. With this aim in view we are carrying out a complete rewrite of the software by using Criterion Renderware. Several very interesting pieces of software such as Open GL and Java 3D have recently been introduced into the video games industry in order to provide the highest level of gaming possible to users. At the same time as we are improving software we have begun experiments with a display device (an Elumens Vision Station® hemispherical projection screen) to which an LCD tablet is added.

The hemisphere will be used to display information in the background of the 3D screen, whilst the LCD tablet will allow the user to interact with the documents in the foreground (for annotations, for example). First of all, we are investigating which scenario to use in order to manage the change in context between the two devices (see [BAU 01]).

Comparing the reading habits of the users of the CNUM site with those who used the 3D workshop will be our next task. We want to link quantitative methods, such as those mentioned earlier, with interviews from those users who read a lot of documents on the CNUM site. In addition to the Bibusages study mentioned earlier, you will also find a methodological review of this type of study in [BRY 00]. The progressive popularization of 3D hardware and of high-speed Internet connections gives us hope that access to digital libraries will become "as easy as a game" [CHR 02] in the near future. Before that though, the processes which have been suggested by different teams contributing to this theme of research must be compared.

Although at first sight it is possible to believe that "the migration of library material to an online environment using the same paper distribution [is] a simple technical feasibility issue" [STE 99], from our particular experience we can conclude that this subject raises more interesting questions for the research community within the information sciences. In their current state, digitized libraries seem to be going through a process of change. Care is taken in the digitization of texts, but not enough care is taken as far as the book itself is concerned, its physical environment (the library), as well as to how the reader treats the book. 3D real time technology enables us to create books from one-dimensional texts, and this idea could also increase communication between libraries and their users.

8.6. Bibliography

[ASS 03] ASSADI H., BEAUVISAGE T. *et al.*, "Usages des bibliothèques en lignes. Projet BibUsages. Rapport final", *France Télécom R&D et Bibliothèque Nationale de France*, July 2003, 58 p. http://www.bnf.fr.

[BAU 01] BAUDISH P., GOOD N., STEWART P., "Focus plus context screens: combining display technology with visualization techniques", *Proc. of UIST'01*, Orlando, United States, November 2001.

[BRO 00] BROWN M.S., SEALES W.B., "Beyond 2D images: Effective 3D imaging for library materials", *Proc. of the 5th ACM Conf. on Digital Libraries*, San Antonio, June 2000, p. 27-36, 2000.

[BRO 01] BROWN M.S., SEALES W.B., "Document restauration using 3D shape: A general Deskewing algorithm for arbitrarily warped documents", *Proc. Int'l Conference on Computer Vision (ICCV'01)*, Vancouver, vol. 2, p. 367-375, 7th-14th July 2001.

[BRO 03] BROWN M.S., TSOI D., "Correcting common distortions in camera-imaged library materials", *Proc. IEEE & ACM Joint Conference on Digital Libraries (JCDL'03)*, Houston, p. 367-368, May 2003.

[BRY 00] BRYAN-KINNS N., BLANDFORD A., "A survey of user studies for digital libraries", *RIDL working paper*, July 2000.

[CAL 98] CALABRETTO S., PINON J.M., BOZZI A., "BAMBI: Système de gestion de manuscrits anciens pour historiens", *Online Document*, vol. 2, no. 3-4, p. 31-50, 1998.

[CAR 96] CARD S.K., ROBERTSON G., YORK W., "The webBook and the web forager: An information workspace for the world-wide-web", *Proc. of ACM CHI'96*, Vancouver, Canada, April 1996.

[CAR 98] CARPENTER L., SHAW S., PRESCOTT A., (eds.), *Towards the Digital Library: The British Library Initiatives for Access Programme*, London, British Lib., p. 61 onwards, 1998.

[CAR 99] CARD S.K., MACKINLAY J.D., SHNEIDERMAN B. (eds.), *Readings in Information Visualization. Using Vision to Think*, San Francisco, Morgan Kaufmann, 1999.

[CAR 04] CARD S.K., HONG L.J.D., MacKinlay E.H., Chi, "3Book: A Scalable 3D Virtual Book", *Proc. of ACM CHI'04*, Austria, April 2004.

[CHE 99] CHEN C., *Information Visualisation and Virtual Environments*, Springer Verlag, 1999.

[CHE 02] CHEN C., BÖRNER, K., "Top ten problems in visual interfaces to digital libraries", in K. Börner, C. Chen (eds.), *Visual Interfaces to Digital Libraries*, LNCS 2539, 2002.

[CHR 02] CHRISTOFFEL M., SCHMITT B., "Accessing libraries as easy as a game", in K. Börner, C. Chen (eds.), *Visual Interfaces to Digital Libraries*, LNCS 2539, 2002.

[CHU 04] CHU Y.C., BAINBRIDGE JONES D., WITTEN I., "Realistic books: A bizarre hommage to an obsolete medium?", *Proc. of ACM-IEEE JCDL'04*, Tucson, June 2004.

[CUB 98a] CUBAUD P., GIRARD D., "ABU: Une bibliothèque numérique et son public", *Online Document*, vol. 2(3-4), p. 13-30, 1998.

[CUB 98b] CUBAUD P., THIRIA C., TOPOL A., "Experimenting a 3D Interface for the access to a Digital Library", *Proc. of ACM DL'98*, Pittsburg, United States, June 1998.

[CUB 01] CUBAUD P., TOPOL A., "A VRML-based user interface for an online digitalized antiquarian collection", *Proc. ACM SIGGRAPH Web3D Symp.*, Paderborn, p. 51-59, April 2001.

[CUB 02] CUBAUD P., STOKOWSKI P., TOPOL A., "Mixing browsing and reading activities in a 3D digitalized library", *Proc. of ACM-IEEE JCDL'02*, Portland, United States, June 2002.

[CUB 03a] CUBAUD P., DEBLOCK G., "Le conservatoire numérique des arts et métiers: Historique du projet et organisation du site", *Proceedings from the Conference on the Digitization of Texts and Images, Techniques and Creations*, University of Lille 3, January 2003.

[CUB 03b] CUBAUD P., DUPIRE J., TOPOL A., "Textes, images, volumes: Les bibliothèques numériques au CNAM", *Actes Premières Journées Francophones de la Toile* (JFT), Tours, France, July 2003.

[CUB 04] CUBAUD P., HAAS J.F., TOPOL A., "Numérisation 3D de documents par photogrammétrie", *Proc. 8th Conference on Written Work and Documents in the French-speaking World (CIFED'04)*, La Rochelle, France, June 2004.

[DEB 02] DIGITAL ACCESS TO BOOKS OF THE RENAISSANCE (DEBORA), EC Telematics app. program 75608. http://debora.enssib.fr, 2002.

[DUP 03] DUPIRE J., "Digital libraries at CNAM: 1993-2003", *Proc. of ECDL'03*, Trondheim, Norway, 2003.

[EGE 01] EGELS Y., KASSER M., *Photogrammétrie numérique*, Hermès Science Publications, 2001.

[FIN 00] FINGERHUT M., "Le site web de la bibliothèque considéré comme un espace", *Newsletter of the Libraries of France*, vol. 45, no. 3, May 2000.

[FOX 97] FOX E., KIPP N. *et al.*, "Networked digital library of theses and dissertations: An international effort unlocking university resources", *D-lib magazine*, September 1997.

[LEC 98] LECOLINET E., LIKFORMAN L. *et al.*, "An integrated reading and editing environment for scholarly research on literary works and their handwritten sources", *Proc. ACM Conf. on Digital Libraries*, Pittsburgh, p. 144-151, 1998.

[MAN 98] MANGUEL A., *Une histoire de la lecture*, Actes Sud, 1998.

[OHA 97] O'HARA K., SELLEN A., "A comparison of reading paper and on-line documents", *Proc. ACM Conf. on Human Factors in Comp. Sys. (CHI'97)*, Atlanta, 1997.

[OTL 34] OTLET P., *Traité de Documentation*, Le livre sur le livre, théorie et pratique, Bruxelle, van Keenberghen, 1934, Liège re-edition, CLPCF, p. 238 and p. 289 onwards 1989.

[PIL 01] PILU M., "Undoing paper curl distortion using applicable surfaces", *Proc. Computer Vision and Pattern Recognition (CVPR 01)*, Hawaii, 2001.

[PLE 01] PLÉNACOSTE P., LECOLINET E. *et al*, "Zoomable and 3D representations for digital libraries", *Proc of IHM-HCI 2001*, Lille, September 2001.

[RAU 00] RAUBER A., BINA H., "Visualizing electronic document repositories: Drawing books and papers in a digital library", *Proc. of the 5th IFIP 2.6 Work. Conf. on Visual Databases Systems (VDB5)*, Fukuoka, Japan, May 2000.

[ROB 93] ROBERTSON G., CARD S., MACKINLAY J., "Information visualization using 3D interactive visualization", *Comm. of the ACM*, 36(4), p. 56-71, April 1993.

[ROZ 01] ROZET B., DEBLOCK G., CUBAUD P., "Le conservatoire numérique des arts et métiers: Une création partenariale", *Newsletter of the Libraries of France*, vol. 46, no. 4, p. 43-49, 2001.

[STE 99] STERN D. (eds.), "Digital libraries. Philosophies, technical design considerations, and example scenarios", *Science and Tech. Libraries*, 17 (3-4), 1999.

[TOP 02] TOPOL A., Interaction 3D pour les paysages informationnels, Doctoral thesis in informatics, National Conservatory of Arts and Crafts, December 2002.

[WAD 95] WADA T., UKIDA H., Matsuyama T., "Shape from shading with interreflections under proximal light source: 3D shape reconstruction of unfolded book surface from a scanner image", *Proc. International Conference of Computer Vision (ICCV'95)*, p. 66-71, 1995.

[WIT 03] WITTEN I. *et al.*, "How to turn the page", *Proc. IEEE & ACM Joint Conference on Digital Libraries (JCDL'03)*, p. 186-188, Houston, May 2003.

This page intentionally left blank

Chapter 9

Using Facets to Classify and Access Digital Resources: Proposal and Example

9.1. Introduction

The Internet and the World Wide Web have cleared new paths toward general and specialized information, now created and transferred into a digital format. Since it facilitates information sharing and exchange between field specialists, the Internet exerts an influence over the development of knowledge fields as well as on the evolution of scientific culture. Yet, relevant sources of information remain under-used by those to whom they would be most useful, because they were not retrieved at the exact moment they were needed. Information specialists know that making information available on the Web is not equivalent to making it fully accessible and usable. Pertinent information only becomes truly accessible when it has been described, indexed and classified; better retrieval results are obtained when information searchers are given the opportunity to search a reservoir of resources using assigned keywords, a classification structure or, even better, a combination of both.

Significant amounts of education-related information sources currently reside on the Web. Practitioners and researchers in education can access this information through more or less efficient keyword searches, by browsing general directories such as *Yahoo!* (www.yahoo.com), *La Toile du Québec* (www.toile.qc.ca), etc., or by exploring virtual special collections such as those maintained in the *Gateway to Educational Materials* or by the Australian Education Network (EdNA). These

virtual libraries offer classified collections of live links to relevant resources that have been evaluated, selected, and described by specialists. Even if we do not yet have much statistical data relating to methods and frequency of usage, the development and management of virtual libraries have become established practices.

Many information specialists are interested in virtual collections and in the classification structures, traditional or innovative, that facilitate their exploration and exploitation. Ad hoc web classification structures are described as intuitive and user-friendly by their designers. However, a good number of them share significant weaknesses due to their lack of concern for any form of standardization, and to their disregard for known theoretical principles of knowledge organization such as facet analysis, class citation order, etc. On the other hand, these structures do exhibit several characteristics deemed essential in good and efficient classification structures. Common to lists of desirable qualities proposed by specialists are the following: simplicity, logic, flexibility, hospitality, authority and specificity [IYE 95, MOL 95]

In previous papers [HUD 00, HUD 03], we proposed general observations on home-grown classification structures used to organize education-related sources in general web directories and in special web-based libraries, with an emphasis on macro-levels of subdivision. In the framework of a 2003-2006 project funded by the Fond québécois pour la recherche sur la société et la culture (FQRSC)[1], we were able to considerably refine our analysis of these structures with the objective of better identifying their qualities and defects, their strengths and weaknesses [HUD 05]. The results of this micro-analysis were taken into account in the second part of the project, which involved the development of a faceted classification structure to be used for organizing and accessing collections of web resources of interest to Francophone specialists and researchers in the field of education.

In this chapter, we first summarize the results of our analysis of existing classification structures. Then, we introduce our proposal for an alternate, less complex classification and access structure, supported by generic facets, to be implemented in a collection of virtual resources in education.

9.2. Examining existing classification structures

To complete a multidimensional description of the characteristics of classification structures currently used to organize virtual collections of web resources in education, we applied a model developed by Sabine Mas, Doctoral candidate at the University of Montreal, as part of her work on the classification of electronic records residing on personal workstations in large organizations. The model, which was strongly dependent on the theory of knowledge organization, led us to an examination of the web-based classification schemes' structure, logic and semantics. A complete description of the model can be found in [HUD 05].

9.2.1. *Sample*

A sample of five web-based education libraries was established, to which was added the education class of a virtual library covering all social sciences. Only those libraries with hierarchically organized collections were considered for this sample.

The following virtual collections were eventually retained: *The Educator's Reference Desk* (www.eduref.org); *EdNA* (www.edna.edu.au/edna); *INTUTE Education* (www.intute.ac.uk/socialsciences/education); *The Gateway to Educational Materials* (thegateway.org); *Education Index* (www.educationindex.com); and the *Education Virtual Library* (www.csu.edu.au/education/library.html). We would have liked to include in this sample a virtual library offering resources and access in the French language, but none of the sites visited at the time were judged appropriate for our purpose. However, since the majority of our observations are quite general in nature, we have no reason to believe that they would not apply also to virtual collections of resources in French, Spanish, etc. All the libraries in our sample were visited several times in the months of January to March 2005; classification structures, however, were imported for analysis at a single point in time.

9.2.2. *Methodology*

Each classification structure was collapsed and transferred to a standard Windows file directory format. This transfer was necessary for the structural analysis, the largest part of which would be achieved with the help of the *PDS* (*Personal Document Space*) software, used by authorization of its designers [GON 03].

Excel-based tables were produced during phases 2 and 3 of the analysis, which involved comparison and interpretation of qualitative data relating to the logical and semantic characteristics of each classification structure. Coding was carried out by

two doctoral candidates working independently of one another, to increase the reliability and validity of results.

9.2.3. *Results and discussion*

Each structure was examined from three essential dimensions: structure, logic, and semantics. Here, we present a summary of results and related observations; a complete description of results and interpretations for each stage of the analysis will be found in [HUD 05] and [HUD 06]. The following abbreviations have been used when referring to the six structures that were examined: *ERD* (*Educator's Reference Desk*), *EdNA* (*Australian Education Network*), *INT* (*INTUTE Education), GEM* (*Gateway to Educational Materials*), *EI* (*Education Index*) and *EVL* (*Education Virtual Library*).

9.2.3.1. *Structure*

The structural analysis generated sets of quantitative data relating to maximum, minimum and average numbers of classes and hierarchical levels, as well as data on the branching factor or average number of classes at each level.

The data revealed little more than what we already knew, or at least suspected. The average of 7.83 main classes in all structures is well below the minimum of 10 top classes judged efficient for organizing resources in a specialized field and, not surprisingly, the higher numbers of distinct main classes are found in the deeper and/or most complex structures (*ERD*, *EdNA*). There is a clear distinction in the total number of distinct classes between the four hierarchical structures (*ERD*, *EdNA*, *INT*, *GEM*) and the two faceted structures (*EI*, *EVL*), the former being significantly more developed. The average number of hierarchical levels, at 3.33, corresponds to a *de facto* standard number of levels recommended and common for general web classification structures; it is generally assumed that the majority of information seekers will navigate only to the third or fourth level of division, each level corresponding to a mouse click, before reorienting their search. Faceted structures (*EI* and *EVL*) are less balanced than their hierarchical counterparts, starting with a narrow choice of top classes (two and four respectively), and then quickly expanding their semantic coverage by way of long alphabetical lists of sub-classes at the second level. None of the six structures is overly complex. This obvious lack of specificity will allow for no more than a broad classification of resources in the virtual collection.

9.2.3.2. *Logic*

Qualitative data relating to the logic dimension of each classification structure was obtained through manual examination and interpretation; three sets of data are available.

The first set of data describes the dividing criteria that are applied at the first three levels of the structure. Six potential values were selected from a list of eight described by [ZIN 02]. They are: Subject, Object, Target audience, Format (or External form), Reference (or Internal form) and Location. A second set of data describes the nature of the relation linking classes at the top three levels of each classification structure. Potential values are: 1) the generic relationship, where the lower level class is a specific type of the object, event, etc. named at the higher level; 2) the partitive relationship, where the lower level class is a component of the object, event, etc. named at the higher level; 3) the instance relationship, where the lower level class is a particular object, event, etc. serving as an example of the object, event, etc. named at the higher level; 4) the contextual relationship, where higher and lower classes are found in the same environment but not in the same natural or logical taxonomy. The third set of data describes the internal arrangement of classes at the top three levels of the structure; three cases are observed: the alphabetical arrangement, the systematic arrangement, and the "mixed" arrangement.

The practice of mixing various principles of division in a developing hierarchy is contrary to theoretical principles of classification because it creates classes of resources that are not mutually exclusive, thus "causing uncertainty for the browser when he has to select a category" [VAN 98, p. 382]. Nevertheless, such a mix is found at the top level of all structures in our sample. In *ERD* for example, Subject, Format, and Target audience are at the root of first level classes. However, whether classes are mutually exclusive or not may not be a problem, providing resources can be assigned to more than one class at the same hierarchical level. This possibility remains to be verified; we do not know at this time if structures and processes now allow for greater flexibility in classification than was possible in traditional, pre-Internet contexts.

Objects, events, etc. are most frequently linked through hierarchical relations of a contextual nature; this is also the case in bibliographic classifications such as the *Dewey Decimal Classification (DDC)* or the *Universal Decimal Classification (UDC)*. This choice of a contextual (e.g. Educational management ►Educational facilities) rather than a truly generic relationship (e.g. Educational institutions ►Secondary schools) contributes to making the classification structure more hospitable and capable of integrating easily new classes and specific topics. The simple and familiar alphabetical display of classes is also beneficial to the

hospitality of the structure, and undoubtedly preferable to a more or less obscure systematic arrangement reflecting the designer's personal view of the world.

The objectives of this project did not include comparing innovative with traditional structures on the basis of logic, but enough is known about the *DDC* and the *UDC* to suggest that our sample ad hoc structures are neither more difficult nor easier to navigate than traditional structures, which are generally considered complex and not user-friendly.

9.2.3.3. *Semantics*

The semantic analysis provided data on conceptual and terminological concordance of our sample classification structures with authoritative sources in the field of education. Results were obtained through a standard methodology for establishing compatibility, involving manual examination of data and the coder's judgment as to degree of concordance. Possible values were: Full or Partial terminological concordance, Full or Partial conceptual concordance, and No concordance.

We compared *ERD*'s class denominations at the top three levels with those appearing in the table of contents of an authoritative reference tool, the *Encyclopaedia of Educational Research*, 6th edition. Terminological and conceptual concordance remains rather low, with the latter being slightly higher, as could be expected.

Top level class denominations in *ERD* were also compared to captions in the web version of the *Dewey Decimal Classification* (*DDC*). Surprisingly, we note that concordance between *ERD* and *DDC* is higher than concordance between *ERD* and the specialized reference work. This may be explained by the fact that the Dewey system is already used for classifying millions of documents and subjects; most likely, this contributes to making it close to being conceptually complete at the first five or six levels of hierarchy, even in specialized areas. The comparison with Dewey also benefits from the encyclopaedic character of its coverage; when a concept is only peripherally related to education, it will not be found in a specialized reference tool, but it is likely to be found somewhere in a general knowledge organization structure.

Our results reveal that partial concordance is always higher than full concordance, at both conceptual and terminological levels. A quick review of top level classes shows that a single facet, Educational levels, is present in all six structures, either in the form of a first-level inclusive class or as a list of constituting categories (Primary education, Secondary education, etc.). This is not surprising, given that there is not a single (or a best) way to segment and organize the world of

concepts, even within the same cultural, political, disciplinary, etc. context. This particularity increases slightly the complexity of the structure, without affecting its authority.

9.2.3.4. *Conclusion*

This first part of our research project confirmed that the hierarchical model remains popular for organizing web resources in specialized virtual collections. In our sample, hierarchies were contextual rather than generic, not overly complex and not very specific. Choice, arrangement and sequence of classes within the structures appeared logical enough to make them easy to apprehend and navigate. However, we observed that the structures were not very flexible and did not appear to benefit much from the technological environment in which they had developed and were now applied.

Among the six structures that were examined, two appeared closer to an alternate model for organizing objects, subjects, and classes: *Education Index* (www.educationindex.com) and *Education Virtual Library* (www.csu.edu.au/education/library.html) made use of explicit facets. However, these two structures were the least developed and the least balanced of all, and it was not possible to extrapolate on the usefulness of facets to structure and access virtual collections. It is this alternative faceted model that we have explored in the second part of our research project.

9.3. A faceted structure to organize and access resources in a virtual library in education

The use of faceted structures to organize and access specialized digital resources is not yet widespread, even if it seems obvious that contemporary networks constitute the ideal environment for implementing the analytico-synthetic principles and practices suggested by S.R. Ranganathan in the 1930s. The facet is a characteristic, an indicator, a criterion that may be used to subdivide a class or a set of objects in homogenous subsets. Age, gender or place of residence, for example, are facets that can be used to create subsets of persons. The same facets can be used to identify fairly precisely each member of the original group: thus, X is a woman, belonging to the 30-39 age group, residing in the Montérégie administrative region of Quebec, etc. A facet may be usable with any group of objects or subjects (for example, agent, process, property, method), or apply only to certain categories of concepts and objects or within a single discipline (for example, educational level or source of financing, in education).

When a choice has been made to work with a faceted structure rather than with a strictly hierarchical structure, classifying an information resource no longer consists of locating its main subject on a pre-drawn thematic map; rather, it requires a complete analysis of the subject using in turn all facets or perspectives from which it can be considered. The subject can then be represented very precisely. Intellectually, the faceted classification offers several benefits:

1) Starting with a much smaller number of distinct classes, it authorizes a much more refined representation of many subjects than enumerative classifications (such as the *DDC* or the *UDC*) do.

2) It is more flexible and adapts easily to conceptual evolution and renewal; it is always possible to modify isolates or values attached to a facet, or even to add a facet, without affecting to a large extent the global structure of the system.

3) When explicit facets are used to organize and access subjects and collections, it becomes possible to optimize automated search strategies since a subject may then be retrieved using any one of the facets that has been used to describe it.

Several knowledge organization specialists have shown a definite interest in facets in their discussion of organization and access to web resources [BRO 02, ELL 00, LAB 06]. Van der Walt [VAN 04] and Zins and Guttmann [ZIN 00] have designed faceted structures to describe and classify specific domains; we used these structures as examples in designing our own proposal for an alternative to a strictly hierarchical structuring of virtual collections in education.

Our structure was developed in several stages, using a deductive approach strongly dependent on literary warrant: 1) creation of a sample virtual collection of web resources in education; 2) classification of each resource using a traditional classification scheme (*DDC*), as well as the structure used by *The Educator's Reference Desk (ERD)*; 3) indexing of each resource using a traditional thesaurus: *EDUthès: Thésaurus de l'éducation* (http://www.cdc.qc.ca/eduthes.html), and design of a bank of candidate descriptors and potential isolates; 4) identification of structural facets needed for content analysis and representation; 5) construction of the faceted structure.

9.3.1. *Creating a special virtual collection of web resources in education*

The identification of relevant facets in our chosen discipline, the development of a bank of descriptors and the planned testing of the proposed classification structure all made it necessary to establish a small virtual collection of web resources in education, which would exhibit a wide variety of contents and types of documents.

Our virtual collection grew over a period of several weeks through exploration of the Web with its main search engines, and by targeted inspection of sites created and maintained by Faculties of Education in Francophone universities, by government departments and agencies, and by national and international associations with ties to education. Following links to related resources proposed by selected sites also proved productive.

25 themes or formal criteria that had been suggested by a survey a existing virtual libraries in education, by examining the detailed summary of the *Encyclopaedia of Educational Research*, as well as by a literature review on information needs and behavior of professors of education served as departure points for exploring the Web. Our objective of building a collection of 400 distinct resources was achieved at the price of long hours of wandering on networks, searching for document types and for contents not yet represented in our sample. We were ultimately able to work with a collection of 408 resources (or live links). Table 9.1 describes the contents of this collection.

Theme	*N* resources	Theme	*N* resources
Legislation	68	Teaching	46
Research	44	Statistics	38
Theory/Theoreticians	36	System and reform	26
Policy	21	Career	18
Equity/Discrimination	13	Philosophy of educ.	11
History of education	9	Terminology	9
Academic publishing	9	Theses	8
Information management	7	General management	6
Associations	6	Economy of education	6
Psychology	6	Comparative educ.	5
Standards	4	Ethics	4
Unions and labor issues	4	Legal issues	3
Leadership	1	TOTAL	408

Table 9.1. *Contents of the sample virtual collection*

Our sample collection was composed of a good variety of document types, from portals to journal articles available in pdf format, from catalogs and bibliographies to institutional websites, from reference works to sponsored expert reports. A brief description of each resource was created at the time of selection; the description included as a minimum a significant title, a precise URL, an indication of source, and a statement as to the availability of links to related resources (see Figure 9.1).

Digital collection development, itself a growing and active field of research, is of course far from easy, if only because of the difficulty of determining what exactly a web document is. At this stage of our project, we chose the most inclusive definition of the web document (web page, websites, series, etc.) and did not concern ourselves with any potential value that could be assigned to a particular resource by an information specialist, to discriminate, for example, between a personal page and an institutional page. In the case of series (weekly, monthly, etc.), titles were included once only and the corresponding URL led to a resource bearing a specific date. It did not appear necessary, for the purpose of this project, to apply cataloging rules relating to title modifications, even if this may have led to the discovery of additional resources.

9.3.2. *Classification and indexing*

Each one of the 408 web resources selected for our sample collection was classified using the *DDC* and the hierarchical structure proposed by the *ERD*. The choice of *DDC* was justified by the significant presence of this classification scheme in a wide variety of information environments (including the World Wide Web) and by our own confirmation of the semantic relevance of this traditional structure for collections of resources in education [HUD 05]. *ERD* is the special virtual library whose access structure was considered the most efficient of all that were evaluated in the first part of our research project.

The main objective of the classification operation, at times very complex as we can imagine, was the identification of structural facets [MAN 99] applicable to the field of education; such process has been recommended by Vickery and the Classification Research Group (CRG) [VIC 60]. Furthermore, classification with these systems would later make possible a comparison of three quite different organizing and access structures, on the basis of efficiency.

Each resource was also assigned a set of descriptors proposed in *ÉDUthès: Thésaurus de l'éducation. ÉDUthès*, created and maintained in Quebec, offers a list of close to 4,000 controlled terms describing major subject areas in the field.

Figure 9.1 shows a descriptive record following selection, description, classification and indexing.

Title	*Equality in European educational systems: a set of indicators*
URL	http://europa.eu.int/comm/education/programmes/socrates/observation/equality_fr.pdf
Theme	Equal educational questions ?
Source	European Group on equal educational systems /Project supported by European Commission, Education and Culture Direction
Document type	Text document
Links	No
DDC	370.94 Education – Geographical treatment – Europe 379.26 Educational equalization (Equal educational opportunity)
ERD	General Education – Comparative Education/ Educational Management – No Child Left Behind
ÉDUthès	Comparative Education/School system /Social Indicator/Economic Indicator/Output indicator/ Social Differences/Educational equalization
Abstract	Final report exploring unequal educational factors, expressing them in indicators (motivation, social inequality, cultural difference, etc.)

Figure 9.1. *Example of descriptive record*

Once the tasks of classification and indexing had been completed, we had at our disposal a reservoir of descriptors which could later be used as precise values, or isolates, in the faceted structure under development.

9.3.3. *Development of a faceted classification structure*

We chose to create a faceted structure rather than a traditional, strictly hierarchical structure, because of the representational flexibility made possible by the use of facets. As our structure was expanding, we kept in mind and defined our objectives in accordance with desirable characteristics of a classification scheme used with digital resources: simplicity, logic, flexibility, hospitality, authority and specificity [MOL 95].

9.3.3.1. *Choosing facets*

The information acquired during previous stages of our research project, as well as the set of titles and descriptors that were already available, allowed us to suggest that a minimum of five generic facets were needed to describe, structure and access a virtual collection of resources in education. These facets are: AGENT (who?), ACTIVITY or process (what?), METHOD or tool (how?), SPACE or context (where?) and TIME (when?). In order to describe and access documents offering more general information not linked directly to any of the preceding questions, a supplementary facet, FOUNDATIONS, was added, as exemplified in the model proposed by Zins and Guttmann [ZIN 00]. The document type facet was excluded from the structure; we consider that this information is of a bibliographic nature and believe that if it became necessary, a straight list of document formats could be offered to the information searcher wanting to use this criterion to restrict and filter a set of retrieved resources.

The proposed classification structure was developed using, as often as it was possible to do so, a principle of division by essential characteristic rather than the principle of contextual hierarchy applied to most contemporary classification structures. In education, there exist two types of AGENTS: Persons/Individuals and Organizations; subsets of persons may be created on the basis of gender, age, and role or function. Teaching, an ACTIVITY, can be described from the perspective of target user (<user>), of objectives (<goals and objectives>), of level (<educational level>), of subject being taught (<discipline/subject>), of method used to do so (<method>).

It was obviously not always possible to respect the principle of division by essential characteristic, and we were at times forced to resort instead to a more traditional thematic division, even at the highest levels within the structure. This was the case for the FOUNDATIONS facet (Table 9.2).

	Education (introduction)
FOUNDATIONS	History of Education
	Philosophy of Education
	Theory of Education

Table 9.2. *First level division in the FOUNDATIONS facet*

To preserve ease of access and user-friendliness, it was decided that the structure would not develop beyond the fifth level of division, that the total number of isolates could not be larger than 400, and that the number of isolates at the deepest level of the structure would be as balanced as possible. Everywhere, except for the top facets, alphabetical order was preferred to systematic order of classes.

9.3.3.2. *Naming classes and choosing isolates*

Each distinct class in a faceted structure bears a specific name and is referred to as an "isolate". Names are normally those that are found most frequently in the literature of a discipline and in its major reference sources. Our bank of potential isolates was established on the basis of keywords and terms used in the titles and summaries of resources in our virtual collection, of terms found in corresponding captions in the *DDC,* of terms used in *ERD* where the first three levels of the hierarchy had been translated in French by a member of our team, and finally of *EDUthès* descriptors. We must specify that semantic control of terms appearing in our faceted structure is minimal; the chosen term is the most current, the simplest in form, and the most accessible; grammatical gender and number have been standardized, but equivalence with synonyms, homonyms or homographs has not been documented. Control of the phenomenon of conceptual equivalence could be achieved in an index, but could not be made explicit within the structure itself without decreasing its ease of use and user-friendliness. Since the classification structure should normally be used by researchers and discipline specialists, we believe that the lack of complete semantic control (as would be found, for example, in a traditional thesaurus) should not be a source of major problems.

Table 9.3 expands the Education class of the ACTIVITY facet, and lists a series of isolates appearing at the fourth level of the structure.

Education		Adults
	by user	Immigrants
		Disabled persons
	by discipline/subject	Administration
		Arts and Literature
		Languages
		Mathematics
		Applied Sciences
		Educational Sciences
		Nature Sciences
		Life Sciences

		Humanities and Social Sciences
	by goals and objectives	Literary studies
		Civic education
		Social education
		Basic education
		Lifelong education
		Higher training
		Vocational training
		Technical training
	by method	E-learning
		Corrective education
		Education by team
		Holistic education
		Authoritative education
		Education by immersion
	by level	Preschool education
		Primary education
		Secondary education
		Post-secondary education

Table 9.3. *Expansion of the Education class in the ACTIVITY facet*

The classes thus created, whose name often results from a process of pre-coordinating concepts of different nature (as in Teaching + Disabled persons or Teaching + Humanities and Social sciences), are not mutually exclusive. Assignment of a resource to multiple classes must then be not only authorized but also strongly recommended, so that several access paths to a relevant resource are created. It should also be noted that the lists of isolates are far from exhaustive; in a faceted structure, this is not a problem since the structure itself allows for new subjects or set of subjects, expressed in the form of isolates, to be added without making it necessary to review what has already been established and validated.

At the time of planning a first application of the newly developed structure to describe and organize our virtual collection of 408 resources, the structure was composed of:

– 6 top level facets representing as many departure points for navigating toward specific subjects;

– 25 second-level classes, 16 of which are not an expression of a principle of division and can be used to describe resources (for example: ACTIVITY ►Communication);

– 86 third-level classes, 70 of which are not an expression of a principle of division and can be used to describe resources (for example: TIME ►"*socio-cultural time*" ►School year);

– 142 classes at the fourth level, among which are two expressions of a principle of division and cannot be used to describe resources (for example: SPACE ►Institutional space ►Teaching institutions ►"*by level*");

– 62 classes at the fifth and deepest level of the structure, all usable to describe resources (for example: AGENT ►"*persons*" ►"*by role or function*" ►Administrator ►Registrar).

288 isolates, judged of interest to researchers and specialists of education, were available in the structure that was used to organize the resources of the virtual sample collection.

9.3.4. *Using the faceted structure*

The developing faceted structure was used to describe and organize the 408 specialized resources in our sample virtual collection in education. Although the lack of the most appropriate technological interface did not allow for optimal exploitation of the structure, it is possible for us to comment at this time on extension and conceptual coverage, navigational logic, flexibility, extensibility and structural hospitality. Our perspective is that of the classifier using the structure to describe and organize a collection; in a later stage of this research, the structure will also be used by information searchers to access the sample collection.

9.3.4.1. *Extension and conceptual coverage*

As would be the case in any indexing and retrieval language constructed *a posteriori* to organize and access an actual collection, conceptual coverage in our faceted structure is necessarily oriented and limited by the contents of the sample collection. We had estimated that 400 isolates would be necessary to describe and organize a basic collection destined to be used by education specialists; this number was not reached. However, the characteristics of extensibility and structural hospitality of faceted structures should allow systematic expansion of the structure to parallel that of the collection, and ultimately that of the field.

9.3.4.2. *Navigational logic*

To make it easier to navigate the structure, the number of top and lower level facets allowing for its expansion was deliberately restricted. When the principle of division by essential characteristic was no longer applicable, generic hierarchy came into play to enable the development of deeper structural levels (Table 9.4).

AGENTS ►*Persons* ►*by role or function* ► Teacher ►Associate Professor
ACTIVITY ►Teaching/Learning ►*by level* ►Post-secondary education ►Academic education
ACTIVITY ►Research ►*by method* ►qualitative research
SPACE ►Institutional space ►Educational institution ► *by financing* ►Private institution
SPACE ►Institutional space ►Educational institution ►*by level* ►Secondary school
METHOD ►*for teaching and learning* ►Educational technology ►Multimedia technology

Table 9.4. *Application of the principle of division by essential characteristic, supplemented by that of generic hierarchy*

To preserve user-friendliness, alphabetical order was applied at all levels of the structure but the top one.

9.3.4.3. *Flexibility*

Flexibility may be assessed from two distinct perspectives, flexibility relates to contents, and it also relates to usage, recommended and actual.

Simplicity and ease of navigation were primary objectives in this exercise. To attain these goals, redundancy in contents became necessary. Such redundancy allows for an isolate, for example Standards, to appear in various places within the structure, possibly under different top level facets, within contexts that provide the represented concept with various, but always appropriate meanings (Table 9.5).

	for assessment	Standard
METHOD	*for management*	Standard
	for research	Standard

Table 9.5. *Example of redundancy by reuse of an isolate in various contexts*

It is in the expansion of the facet METHOD or tool that the largest number of cases of redundancy by reuse of the same isolate in various contexts will be found.

Redundancy is also observed in class names, as in Student ►Student with learning disabilities, or Teaching ►"*by level*" ►Post-secondary teaching ►University teaching. This second type of redundancy, which we will call lexical redundancy, allows for recognition of the meaning given to an isolate, even when it is encountered outside of its structural and navigational context (in an alphabetical index for example).

The flexibility of the faceted structure is also a factor of its use, not only for organizing resources in homogenous groupings, but also for describing each resource in a collection. The process of classifying using a faceted structure is quite similar to that of indexing the contents of a resource. Working with a faceted structure, the classifier cannot simply assign a resource to the generally conceptually complex class to which its main subject belong, this class being identified on the document itself by some symbol. Instead, the classifier must assign a resource to multiple conceptually simple classes to reflect all facets of the subject. Table 9.6 provides examples of multiple allocations of sample resources to various facets and classes within.

Title of resource	Classification
AFIRSE: (Association francophone internationale de recherche scientifique en éducation)	FOUNDATIONS ►Theory of Education ►Disciplinary contributions AGENT ►Organization ►Association ►Scholarly association ACTIVITY ►Research ACTIVITY ►Communication
John Locke (1632-1704) Some Thoughts on Education	FOUNDATIONS ►Philosophy of Education FOUNDATIONS ► Theory of Education ACTIVITY ►Teaching/Learning ►*by objective* ►Civic education SPACE ►Geopolitical space ►Europe ►England TIME ►Chronological time ►Before 19th
Educational program of Quebec School: primary and secondary education	AGENT ►Organization ►Department/ Organization/Government agency ►Department of Education ACTIVITY ►Teaching/Learning ►*by level* ►Primary School ACTIVITY ► Teaching/Learning ► *by level* ►Secondary School METHOD ►*for management* ►Program SPACE ►Geopolitical space ►North America ►Canada ►Quebec
Microsoft Education	AGENT ►Organization ► Non-governmental commercial/organization MEAN ►for teaching and learning ►Educational technology ►Computer sciences technology SPACE ►Geopolitical space ►Europe ►France

Academic Research Information System	AGENT ►Persons ►*by role or function* ►Researcher
	AGENT ►Persons ► *by role or function* ►Teacher
	SPACE ►Institutional Space ►Education Institution ►*by level* ►University
	ACTIVITY ►Research
	METHOD ►*for research* ►research grant
	METHOD ► *for research* ►research agreement
	SPACE ►Geopolitical space ►North America ►Canada ►Quebec
How do ICT change the pedagogical engineering practices of academic teachers?	FOUNDATION ►Theory of Education ►Pedagogy
	AGENT ►Persons ►*by role or function* ► Teacher
	ACTIVITY ►Teaching/Learning ►*by level* ►Post-secondary Education ► Academic Education
	SPACE ► Institutional Space ► Educational Institution ► *by level* ►University
	METHOD ►*for teaching and learning* ►Educational technology
	SPACE ► geopolitical space ► North America ►Canada ►Quebec

Table 9.6. *Examples of application of the faceted structure*

These multiple allocations represent as many distinct and logical ways to navigate toward the set of relevant resources needed to fill a need or satisfy an interest. The process will be even more efficient if the interface and search engine authorize simple combinations of facets and values of the type: AGENT + ACTIVITY + SPACE + METHOD as in Teaching person + Team teaching + Secondary school + Educational technology, or complex combinations of the same. In the latter case, several isolates linked to the same facet are related to one another using suggested pre-defined formulas, or formulas created by the information searcher when the need arises, as in : {Director + Assistant director + Registrar} + {Secondary school + Public institution} + Communication.

9.3.4.4. *Extensibility and structural hospitality*

The first application of the proposed faceted structure to describe and organize our sample virtual collection has already led to the integration of specific isolates and of facets allowing for supplementary levels of division, as well as forcing the restructuring of sections of the structure to improve its overall navigational logic. It is thus possible at this time to confirm the extensibility and hospitality of the proposed structure.

By nature, a faceted structure can be extended at any and all of its levels. It is always tempting to add large number of isolates at the lowest and deepest levels, but this is not in the best interest of the information searcher, then at risk of being offered unsatisfactory retrieval sets due to the small number of relevant resources they contain. The main objective of any classification structure remains the grouping of resources linked through their thematic content, their genre, their format, etc., rather than discovery of one or two relevant resources, already made possible by other search methods and strategies, such as a keyword-based search.

The faceted structure is also characterized by its hospitality. The integration of new isolates at the exact place in the structure where they are needed, and within the context that will give them their full meaning, is always possible without modification to the macrostructure itself. Integration of facets expressing a division on the basis of essential characteristic (<*according to…*>, <*by…*>, <*on the basis of…*>) will necessarily bring about the restructuring of sections of the structure, since a list of isolates, until then presented in alphabetical order, will need to be systematically reorganized on the basis of the new point of view. Numerous structural changes are expected during the initial development phases, but when the structure is not complex and not very deep, the frequency of such adjustments will diminish as the bases settle.

9.3.5. *Next steps*

Our work on an alternate structure usable to organize and access a collection of specialized web resources in the field of education is now entering a phase of multiple validations.

Although we have, on occasions, solicited the expertise of education specialists in the course of developing this proposal for a faceted classification structure, we have until now relied mostly on the content of our virtual collection and on the structures proposed by reference works to name, rank and order facets and concepts. A more scientific process of validation is now required; this validation will be sought from researchers in education, primary target users of a virtual library of French resources in this field; most education specialists surveyed at an early stage of this project have confirmed their interest in such an information source.

Validation of ease of navigation, user-friendliness, extensibility and flexibility of the structure will be obtained by having two or more classifiers, who until now have not had any ties with this project or with the project team, use the faceted structure to describe the resources of our sample virtual library. A second application of the structure is already underway.

The third and final validation process will only be launched when an appropriate technological interface allowing for optimal use of the structure to identify and locate relevant resources becomes available. This validation will take place within a controlled experimental environment, where the structure will be used to access and retrieve relevant resources. Education and information specialists will be asked to complete simple and complex search tasks; this will lead to a validation of the proposed content, and it will allow us to compare the performance and efficiency of the alternate classification structure with those of *DDC* and of *ERD*.

9.4. General conclusion

This three-year research project has led us first to confirm that the organization of web resources in virtual libraries has improved considerably over the past 15 years. Whether universal or specialized in their coverage, the "new" classification structures are user-friendly and truly accessible to information searchers, but they remain rather inflexible and, from a logical point of view, still leave much to be desired.

We now have the technology that should allow us to implement principles of analysis and representation proposed in the first half of the 20th century by the Indian

Ranganathan, and to develop and use classificatory structures that are flexible, hospitable and better adapted to the electronic networked environment.

It is within this stream of thinking that we propose an alternative to the strictly hierarchical organization of specialized web resources, with a view to facilitating access and retrieval. The development of a generic structure based on facets for organizing and accessing web resources of interest to specialists and researchers in education has allowed us to present the interest of this solution for the classifier. The faceted structure is still developing, and various validation processes are, or will soon be, in place. We hope to be able to also confirm in the near future the interest of the solution for the web searcher.

9.5. Bibliography

[BRO 02] BROUGHTON, V. "Facet analytical theory as a basis for knowledge organization tool in a subject portal", in *Challenges in Knowledge Representation and Organization for the 21st Century: Integration of Knowledge Across Boundaries: Proceedings of the 7th International ISKO Conference, 10-13 July 2002, Granada, Spain*, ed. Maria Lopez Huertas, 135-142, Würzburg, Germany, Ergon, 2002.

[ELL 00] ELLIS, D. A. VASCONCELOS. "The Relevance of facet analysis for World Wide Web subject organization and searching", *Journal of Internet Cataloging* 2, 3/4, 2000, 97-113.

[GON 03] GONCALVES, D.J., J.A. JORGE. "Analyzing personal document spaces". *Proceedings of the 10th International Conference on Human-Computer Interaction, 22-27 June 2003, Crete, Greece,* 2003, available from www.inesc-id.pt/pt/indicadores/Ficheiros/942.pdf , accessed 21 March 2007.

[HUD 00] HUDON, M. "Innovation and tradition in knowledge organization schemes on the Internet, or, Finding one's way in the virtual library". In *Dynamism and Stability in Knowledge Organization: Proceedings of the Sixth International ISKO Conference, 10-13 July 2000, Toronto, Canada*, 35-40. Würzburg, Germany, Ergon, 2000.

[HUD 03] HUDON, M. "Subject access to Web resources in the field of education". In *Subject Retrieval in a Networked Environment: Proceedings of the IFLA Satellite Meeting, held in Dublin, OH, 14-16 August 2001 and sponsored by the IFLA Classification and Indexing Section, the IFLA Information Technology Section, and OCLC*, 83-89. Munich, Germany, K.G. Saur, 2003.

[HUD 05] HUDON, M, S. MAS, D. GAZO. "Structure, logic, and semantics in *ad hoc* classification schemes applied to Web-based libraries in the field of Education". *Canadian Journal of Information and Library Sciences/Revue canadienne des sciences de l'information et de bibliothéconomie*, 29, 3, 2005, 265-288.

[HUD 06] HUDON, M. "Structure, logique et sémantique dans l'organisation des ressources du Web en éducation", in *Pratiques et méthodes de classification du savoir à l'heure d'Internet : Colloque d'ISKO-France organisée dans le cadre de SDC – Semaine de la connaissance 2006,* 26 June 2006, Nantes, France.

[IYE 95] IYER, H. *Classificatory Structures: Concepts, Relations and Representation.* Frankfurt, Germany, Indeks, 1995.

[LAB 06] LA BARRE, K. "A Multi-faceted view: use of facet analysis in the practice of Website organization and access", in *Knowledge Organization for a Global Learning Society: Proceedings of the Ninth International ISKO Conference, 4-7 July 2006, Vienna, Austria*, 359-366, Würzburg, Germany, Ergon, 2006.

[MAN 99] MANIEZ, J. "Du bon usage des facettes". *Documentaliste – Sciences de l'information*, 36, 4/5, 1999, 249-262.

[MOL 95] MOLHOLT, P. "Qualities of classification schemes for the information superhighway", *Cataloging & Classification Quarterly*, 21, 2, 1995, 19-22.

[VAN 98] VAN DER WALT, M. "The structure of classification schemes used in Internet search engines", in *Structures and Relations in Knowledge Organization: Proceedings of the Fifth International ISKO Conference, 25-29 August 1998, Lille, France*, 379-387, Würzburg, Germany, Ergon, 1998.

[VAN 04] VAN DER WALT, M. "A classification scheme for the organization of electronic documents in small, medium, and micro enterprises (SMMEs)", *Knowledge Organization*, 31, 1, 2004, 26-38.

[VIC 60] VICKERY, B. *Faceted Classification: A Guide to Construction and Use of Special Schemes*, London, UK, ASLIB, 1960.

[ZIN 00] ZINS. C. and D. Guttmann. "Structuring Web bibliographic resources: an exemplary subject classification scheme", *Knowledge Organization* 27, 3, 2000, 143-159.

[ZIN 02] ZINS, C. "Models for classifying Internet resources", *Knowledge Organization* 29, 1, 2002, 20-28.

This page intentionally left blank

Chapter 10

Digital Libraries: the Publication of Legal Documents Online within the Info-mediation Service

Online legal documentation covers two types of support with very different aims and uses. CD ROMs, which enable users to have access to a complete document on a PC and on the Internet, most often offer a continuous thread of the latest legal information so that the experts working in this field are continually made aware of the advances in substantive law.

Relying on CD ROMs, which require the use of one PC or an access code with limited access time, combines a stock of stored legal information (judgments, laws in force, legal publications or commentaries on the latest legal information) with a themed index which makes researching the archives (by the publisher in question) and researching new developments made available by legal practitioners online a lot easier. In certain cases, an associated search engine will list the set of responses to different searches following a keyword query; this list will then publish a register of all the editorial sources or developments by carrying out a full text search according to the research criteria. The CD ROM, which is updated every month or every three months, offers professionals in the legal profession all the raw data necessary in order to help them with their cases. However, for each branch of law there is a corresponding CD ROM and its limited digital data prevents interaction between the databases.

Chapter written by Fabien GIRARD DE BARROS.

On the other hand, interaction is the key to carrying out legal research on the Web. The security of legal information provided by the Internet more often than not requires users to pay in order to have access to the data, which is explained prior to connecting to such sites. However, by accessing legal information users are guaranteed the most up-to-date literature and services rather than just storage of information. The essence of online digital documentation as well as its advantages reside in its flexibility and adaptability, the latter only being possible if the access to legal information, in a case in which we are interested, is understood by the service providing the information. There are also two types of site which make it possible to access substantive law. Sites with the latest information publish a list of the most up-to-date legal information and provide a detailed commentary of the key decisions made or of adopted bills. The Internet is used for its speed and ease of access. The other type of site provides interactive access to published databases combined with currently used sources and weekly legal news. These sites can also be used as practical tools (as diaries, indexes, rates, lists of sites, judgment orders, etc.) which can help simplify the job of professionals in the legal domain in relation to their declaratory obligations, their memos, or their research on the Internet. The Web is not only used as a media source for broadcasting information which can be made available to the general public in a very short period of time, but also as a logistical support allowing for the interaction of several data sources. This interaction is the result of a search engine which is used in a similar way to the most common servers. Sorting solutions in terms of relevance in relation to researched documents seems to be one of the most valued characteristics of the Internet. This idea of service (secure access to updated legal information in the same portal), which assumes a renewable subscription, is much better than purchasing a CD ROM.

The aim of the online publication is to go beyond the simple information content, to help the documentalist who over time has become info-mediator [NER 04] thanks to new technologies of information and communication, i.e. the documentalist is the key person within the company who:

– searches for the availability, instantaneity and simplicity of information;

– searches for the selection of relevant information in their everyday work;

– tries to share the information with everyone working in the same office.

10.1. Availability, instantaneity and simplicity of information: the minimum requirements for legal publications on the Internet

Being able to navigate through and carry out research on all the information made available to users is proof of how secure legal documents on the Internet are. The number of legal cases in general is increasing. This is why online publishers, when developing the information that they produce and broadcast, take this fact into

account (the increasing number of court cases). Speed, security, exhaustivity and analysis are keywords which documentalists need to understand as well as availability (section 10.1.1), instantaneity (section 10.1.2), and simplicity (section 10.1.3) in the research of legal information.

10.1.1. *Accessing legal information: application of the classic unities of tragedy within the company*

An acceleration in the number of cases, as well as an overproduction in legal, jurisprudential and doctrinal terms, means that documents need to be explained on a daily basis for an increase in productivity. The famous classic unities of tragedy revived by Boileau, i.e. the unities of action, place and time, merit an epitaph on the wall of several companies. When putting these unities of action, place and time into context as far as legal information is concerned, we are talking about the management of a number of computer posts with free or controlled access, as well as the management of subscriptions and the price per square meter of the area dedicated to legal literature.

One of the main characteristics of the Internet is its relative ease of access; from any computer it is possible to access the Internet provided that the user has a subscription with an Internet service provider. Searching through a book or collection of books, as well as the uniqueness of software on CD ROM, slows down the time taken to find a response when looking for a solution requested by the consultant, or even by the company's client in question. There is no need to be reminded of the opportunities that the Internet and intranet provide in terms of consulting the same legal source or legal analysis, modifying the performance as well as the waiting times of consultants (all of which can be done if there are other legal professionals working with the same documents at the same time). Without disturbing the relationship with their documentalist, each legal professional can, from now on, consult all the literature made available to them through the Internet or intranet without having to physically move, whether they are at work or at home.

Furthermore, this implicit sharing of information has forced publishers to develop information as a shared service rather than a unitary service. The increase in number of services available and the increase in the number of people using them simultaneously have meant that the current subscription offers need to be changed. It seems that publishers of legal information who provide subscription services to users who want to access their information are doing much better than those who provide pay as you go services, where the user pays the publisher each time they want to access a document. It lies with the publisher and not the documentalist to find the correct price as regards the information that they have made available online. In order to favor users being able to access the same information at the same

time publishers are already offering multiple connections for users using the same user names and passwords. The publishers can also recognize certain IP addresses which enables the users of these addresses to have direct access to the sites.

By looking at the latest studies relating to productivity in France, it then becomes clear how much information is actually available to users of such websites. Some of them show that despite a significant reduction in working hours France finds itself above a bunch of leading countries with the most productive working time (with an increase of 2.32% in hourly productivity between 1996 and 2002 in comparison to 1.63% between 1990 and 1996). In terms of productivity per hours worked, France comes ahead of both Germany and the USA. One of the reasons given for this improvement lies not only in the rationalization or planning of work, but also in the optimization of information systems. According to [ETT 04], "with the integration of information technology, an increasing share of the workload is not carried out during official working hours but rather during the journey to work, at home and on holiday". The publishers of online legal documents want to make it possible for users of their sites to be able to access the information that they are providing from different places, not just at work or at home.

10.1.2. *Judicial security and the instantaneity of the response*

Urgency is key in the relationship between jurists and documentalists. The drawing up of a contract, of conclusions, or of a simple synthesis forces the legal professional to request the specific services of a documentalist: the documentalist must be certain of the substantive law relating to a case from the raw data that is made available to them in relation to the case in question. Traditionally, doctrinal commentary and legal sources have only been used in paper format, which when trying to find a solution to a particular legal question slows down the research process. The CD ROM has tried to respond to this by using hypertext links, but its limited memory capacity, and the relative lack of improvement of research engines which are generally associated with it, do not make it possible to have an instant response to a given legal question.

Without ignoring the necessary relevance of information found on the Internet, it can be pointed out that the majority of publishers of legal documents intend to work on the interactivity of their collection of documents in order to make them available to subscribers on the Internet. The publishers also want to benefit from the advantages that search engines provide.

The Internet or intranet makes it possible to respond almost instantaneously to a legal question which has been asked in the form of key words only. The publishers of these legal documents, depending on whether they have developed their services

and collection of work on the Internet, will offer a synthesized judicial analysis as well as providing an explanation as to how the legal system in question works and provide a doctrinal commentary related to the judicial analysis and/or the updated version of the analysis. This is what some people call hyper-referencing. The documentalist or the jurist will then find a complete chain of legal information available to them through just one click of the mouse. Each person will be able to work on the support that suits them best (synthesis, source or doctrine) without being forbidden from using another approach for the same legal problem.

However, if instantaneity implies the almost automatic availability of corpus of identified and ordered solutions, it also implies that there is a continual update of the raw or editorial contents. The publishers of legal documents can only follow the frantic rhythm of prescriptive production when it comes to helping the documentalist/info-mediator. The broadcasting of commentaries relating to the latest information in the quickest deadlines possible is no longer sufficient; the databases or published judicial online encyclopedias will need to be consolidated and updated within the quickest possible timeframes. This updating process can only take place as soon as the document is first put online. Publishers who digitize their paper publications in order to display them on the Internet run the risk of a loss in productivity should there be any reforms or jurisprudential changes to their encyclopedias. The encyclopedias which are available on the Internet are designed for a synthetic presentation, normal interactivity and constant updates, all of which make the online publisher the provider of the most unbiased, truthful and most secure information for the documentalist or jurist.

10.1.3. *The simplicity of access: ergonomics – providing a helping hand with the abundance of information available on the Internet*

Intuitively, the documentalist or jurist uses the Internet for its main function: as a provider of information. They know that they will find a multitude of information relating to their query through databases which can either be free to access or which charge users an access fee. When taking into account the abundance of free legal information available the user does not really understand why they should subscribe to a paying service with an online legal publisher.

Apart from the problem of legal security of information found on the net via a search engine, it must be pointed out that governmental websites, and more generally the French Public Service for Legal Broadcasts (Service Public de Diffusion du Droit or SPDD) has favored the access to raw data relating to French law. The French government's official website, Légifrance, or different ministerial reports published on the Internet are the result of sites which enable users to access the raw data of French law. The significant and continual increase in the number of

people visiting the site of the SPDD is the proof of a need for increasing judicial information on the Internet.

What we can say is that it has become too easy to be able to carry out research on these sites, to the extent where it has become unacceptable. On one hand, not all legal information is free to access. Synthesis and doctrinal commentaries which are protected by copyright are rarely available free of charge. On the other hand, the site of the SPDD, which essentially attracts legal professionals, was initially developed for the general public. The SPDD does not apply the required standards when legal professionals access files from the site, far from it. People have been able to access appeal judgments, pending judgments and circulars from the 1980s free of charge. It has been judged that the required standards could be brought into question because the information provided did not result in substantive law, and therefore had to be removed from this service which can be accessed free of charge.

The entire range of online legal publications is present. It is possible to provide a set of raw or published data which is free on certain sites, and which must be paid for on others. The data can either be easily accessed by everyone or can only be accessed by the subscribers of the publication. This is the creation of the one-stop shop for judicial legal publications on the Internet for the use of legal professionals; this is similar to the administrative one-stop shops available to the general public.

The online legal publisher has become a platform for legal literature: they make it much easier to enable access to all aspects of law by offering this multitude of sources and commentaries which have been mentioned earlier in this chapter. Through intuitive ergonomics it is also much easier for people who may only use the sites once to have access to the literature. In offering a type of ergonomics which respects the consultation habits of the legal professionals or documentalists, publishers have progressively succeeded in getting the users of these publications to subscribe to the publications. This is achieved by recreating the users' common computer environment exactly or as closely as possible, i.e. by creating a replica of Windows so that the user is familiar with how to navigate through the online publication. The tools which are needed to replicate a common computer environment are: files, shared-files, billing sections and personalized editing sections. Once these are integrated into the online publications it makes it much easier for the legal professional or the archivist to use such publications.

However, simplicity is not everything. Availability and instantaneity definitely lead to an increase in the volume of information found relating to a particular legal issue. Even if the online judicial publisher will never be able to replace the trained eye of the documentalist in order to determine the effective relevance of information within each legal entity, it is nevertheless their job to organize the information in order to facilitate this selection which is carried out by the documentalist.

10.2. The relevance of information: from the documentalist's know-how to the documentalist/info-mediator

No one understands the legal world better than the documentalist of a company or an institution. It is the documentalist who rewrites interrogations made by colleagues in the legal profession in order to make them more intelligible and also to provide a series of relevant answers which will be of interest to the jurist who is researching information for a particular case. However, the publishers of legal documents on the Internet are releasing information of continually improving quality. It does not take long to find responses to queries if key words are used in the search. As of now the time has come for relevant searches (section 10.2.1) to be carried out in order to receive information in a given context (section 10.2.2). Furthermore, the tasks of the documentalist are centered on their own strengths (finding the relevant information) and by training their colleagues on how to use an online publication tool (section 10.2.3). In order to assume the role of info-mediator, the documentalist's role involves acting as a link between the legal information that is available and the Internet.

10.2.1. *The emergence of relevant search engines*

Online judicial publishers are, above all, computer designers. In effect, the majority of search engines offer searches by reference in integrated texts with Boolean operators. Research is carried out using criteria such as date, period and code. However, certain publishers offer a syntactical and contextual new generation search engine. This search engine gives users the opportunity to directly access official texts using a multi-criteria search on raw data or editorial data; or by a unique and fixed identification number (this reference leads to reliable and direct access for the person carrying out the research).

The Internet user therefore begins their research by using different criteria: key words, dates, titles, names of parties, number of the legal text or decision. The search varies depending on what object is used to carry out the search:

– carrying out a search in raw data sources: this method of searching is used, for example, when searching through jurisdictional decisions, legislative dispositions, or endorsements of a collective agreement, all of which can be accessed in their official versions;

– carrying out a search in the editorial database: as soon as the user enters the editorial database, they express the subject they want to research in a few words. The search engine displays the results corresponding to the words used by the user. The results are displayed in order of relevance defined by the research algorithm. By choosing the legal context that applies to the user, the user can then immediately

access responses which are relevant to them. In the space of a few moments the user will have found lots of material that corresponds to what they were looking for. The user is then able to consider other aspects of the problem that they would not have initially considered. The user then receives a set of precise responses, sometimes with a link which allows them to access reference texts in their complete version.

The relevance of the search engine stems from a single observation: a legal term mentioned in the title of a chapter, in a section of a chapter, a development or a footnote do not have the same relevance value. The documentalist or the jurist will become more familiar with the search engine as it will list, in terms of priority, the sources of sections of the databases in which certain occurrences are quoted the most often, or are underlined in a title. Furthermore, the creation of certain databases on the Internet has allowed for the development of a synergy between traditional legal terms appearing in different contents and the terms which are the most used, which are inserted as hidden words (for example, liquidation surplus and latent capital gain). The creation of these databases has also made it possible to link the most common acronyms with their full written form (LLC – Limited Liability Company). These databases have even made it possible to give the French translation of an English legal expression (golden parachute and parachute doré).

10.2.2. *Contextualization: first steps towards the relevance of information*

With the availability of these relevant search engines the Internet user then needs the suggested responses to be placed in context to make the search easier for them. Online legal publishers have not only made a point of contextualizing their solutions within their collection of documents (in whatever encyclopedia or database the information is found), they have also placed this solution in its legal environment. It is simply a question of positioning the solution found by the search engine in the tree structure of the database or of the encyclopedia to which the Internet user has subscribed.

A cross-search can also lead to a set of solutions being present in several databases or encyclopedias in such a way that it offers the legal professional all the required information for the question that has been asked.

10.2.3. *Providing training for the jurists: reinforcing the link between the jurist and the documentalist*

It may seem unusual that one of the new tasks of the documentalist is to train colleagues in the legal profession on how to use such a documentation tool. Does this mean that we are interfering with their job? The answer to this question is no.

Info-mediation assumes that the documentalists will share their knowledge in order to improve the quality of searches, and to improve profitability in general.

Judicial publishers offer a wide range of platforms for researching information on the Internet. These platforms have several reading and usage levels. Understanding the intuitive ergonomics in relation to relevant research and the contextualization of information is indeed not one of the easiest tasks of documentation. Jurists who have been rigorously trained to use these platforms by their documentalists and/or by the publishers themselves are able to directly search the source or editorial databases in order to obtain the first documents on a subject that they are working on. However, they should not carry out all of the research relating to a particular issue by using such a platform. There is no need to go back to a documentalist in order to find out the latest cost construction index or the latest jurisprudence from the Court of Appeal on the non-competition clause which is included in work contracts.

Now that the documentalist/info-mediator no longer has to carry out research on less important issues, they can concentrate on carrying out research on the views of magistrates or on rights and laws pertaining to a particular case. The documentalist will also be able to carry out a comparative study of different doctrines written by different judicial publishers. The documentalist will be able to use their knowledge of the legal world to the best of their ability in order to research the most relevant and unbiased information for their colleagues.

The transmission of basic knowledge in relation to the use of edited online databases is a certain asset for the rationalization of a documentation service. This transmission of knowledge constitutes the first step of the sharing of information. However, judicial publishers are trying to go even further in the sharing of knowledge through the use of a judicial intranet by using channels in the form of a newsletter.

10.3. The sharing of judicial information: when the judicial publisher becomes the computer technician

Judicial information is referred to as having strong added value. This strong added value comes from the company's experience, from the analysis of individual know-how or from legal consultations. The information received is intended for private use. Judicial publishers link their documentation with that of the company via a closed circuit network, i.e. intranet (section 10.3.1). In addition, it might be interesting for a company or an institution to share some legal know-how with their clients or members. It is a question of introducing a specially designed channel for a given public (section 10.3.2). The amount of legal information and the adaptability

of legal professionals is such that it might be interesting for a documentation service to manage the latest updates by sending out specific alerts or by sending profiled e-mails to subscribers of such sites (section 10.3.3).

10.3.1. *Intranet: the symbiosis of official information and personal doctrines*

Recently, publishers of online legal documents have been creating tailor-made tools which have allowed for the integration of all documentation belonging to the legal professional. This is achieved by using a unique integration and contribution tool which enables documents produced by the publisher and documents produced by the client to be available in the same space (appointments, forms, doctrinal positions, etc.).

These tools mean that customers do not need to:

– possess and maintain powerful and expensive internal services;

– create databases capable of managing masses of legal documents;

– acquire official texts as stated within the documentation service;

– append these documents with links to other documents;

– create a thesaurus of key words to index the chosen documents.

These teams of engineers and jurists, who have become experts in the creation of contents, help their clients in the definition and installation of their information systems right up to the management of a private space by responding to specific requirements and constraints, notably confidentiality.

10.3.2. *Channels: communication within the communication service*

Certain online legal publications available on the market aim to create a link which would give preference to their clients or members. In order to achieve this, a common ground which emphasizes knowledge and know-how needs to be established. One of these categories of information which both legal professionals and clients are fond of is undoubtedly that of legal information. Certain legal publishers are thinking of introducing a regularly updated channel on an external site which is now possible thanks to their experience gained in the networking of legal documentation.

This channel would be developed and broadcast by the company or institution itself; the corresponding web pages would be presented in colored codes, logos and other ways of identifying the subscriber. Regarding the contents of the information

being broadcast, the publisher could choose the content type, the frequency of updates and the publication or writing format for a specific target audience.

These channels, which are tailor-made, offer different legal professionals the possibility of distinguishing themselves from others, as well as marketing their knowledge of new legal systems in order to develop new judicial optimization products.

10.3.3. *The alert and the newsletter: managing updates*

The latest change in the sharing of legal information, the alert and the newsletter, are turning out to be key factors in document updates. Online publishers have been quick to integrate the processing capacities which the world of computing has to offer. The tagging of a set of expressions, terms, or legal terms allows the publisher to trace their history on a day to day basis. This is why publishers are continually giving their subscribers the chance to follow the latest information available in a particular aspect of law, in a particular legal system, or even in a raw data source. The subscribers can follow this latest, most up-to-date information at a pace that suits them.

The day before the release of new or updated versions of a document is no longer the day when publishers try to pre-select and prioritize the information that will be included in the new edition (and all the irritations that this leads to). Instead, the day before releasing a new or updated version of a document enables publishers to provide an exhaustive list of themes which would be of interest to the subscriber.

The result of this could be the appearance of a pictogram on the site of the publisher which would be a warning to the subscriber. This update could also be in the form of an SMS in case of urgency, or more probably in the form of newsletters which would be sent to the subscribers of the legal publication.

For the documentalist/info-mediator it is a question of following the correct update for their colleagues and managing the set of letters that they intend to send to the colleagues. Managing the publication of information through the newsletter is also facilitated by the online publishers. The publishers offer the documentalist tools to help them choose articles or the latest news which interests them, to help compose and manage several publications, and to send newsletters as well as managing the sent and returned files.

10.4. Conclusion

It is clear that digital or virtual judicial libraries are far removed from the simple traditional method of filing, researching and synthesizing of information which exists in documentation services. This online library offers everyone working in the legal documentation service the possibility of widening, rationalizing and managing their services for an improvement in filing, researching and synthesizing information as well as improving the links in the chain of information. The documentalist who is worried about reinforcing their role of mediator with the abundance of information and updates is extremely interested in joining the services of these new online publishers whose range of skills (publication, legal and computing) allows for the adaptation of a tailor-made documentation tool. The documentalist/info-mediator assumes their role and establishes their digital library and the platforms of online documentation services which, through time, will merge with one another to create a better level of interactivity and will complement one another in a more effective way.

10.5. Bibliography

[ETT 04] ETTIGHOFFER D., *Trop c'est trop!*, Editions d'Organisation, 2004.

[NER 04]   NERRIERE   R., "Documentation   et   nouveaux   médias: du   archiviste   à l'infomédiateur", *La LJA – Le magazine*, p. 64, 2004.

Chapter 11

What Scholarly and Pedagogic Material is Available Online for the Virtual User Within French Universities?

11.1. The availability of scholarly and pedagogic material online within French universities: an assessment

Today universities produce, broadcast and negotiate the access to more and more digital resources.

The existence of high quality scholarly and pedagogic material, responding to the diversity of the different target audiences, remains an introductory step which is fundamental to the development of the services and uses that are associated with such content. The rapidly increasing number of virtual users, or rather e-learners, is closely linked to the quality, amount and diversity of material that is available, not only in English but also in all languages spoken throughout the world.

What scholarly and pedagogic material is available in French universities? Here, the term scholarly and pedagogic material refers to all content which is destined for the use of students or teachers and researchers and which has been subject to a process of selection and validation either within the universities themselves, or carried out by an external body.

In this report we will try to provide some answers in support of the results of different studies which have been carried out or which are being carried out, as well

Chapter written by Ghislaine CHARTRON and Marc MINON.

as providing the results of surveys and detailing our observations made throughout the last few years in this field.

11.1.1. *An economic scale that distinguishes three models*

Let us consider the different dynamics at work which have been inspired by the works of Julien Deceuninck [DEC 02], i.e. the three types of publishing, each of which is related to three different economic models. Sometimes it can be quite difficult to distinguish the boundaries between each of these models:

– published material which is available free of charge is increasingly financed by the state because of the public character of this content or because of the uncertainty of the emerging material that is produced by private individuals;

– published material which is registered as part of a market economy for which universities have acquired access rights;

– published material which is developed in the framework of a common economy, the author gives his work away by broadcasting it on different networks. In this case, the process of validation is collective and carried out by the users themselves.

11.1.2. *Published material as public property*

According to the first economic logic, the published material is considered as a product which must be made accessible to everyone (no exclusion and must be of general interest to everyone). The published material is also considered as unsafe and inappropriate for a market economy. Investment by the state therefore makes it possible to produce this material and to insure that it is made available and can be accessed from anywhere [SCH 97].

As far as France is concerned, the development of a public goods economy for material which is published for tertiary education and research has been by no means insignificant over the last few years. In order to see how this has worked, you just need to look at the availability of journal archives whose digitization have been financed by the state (the websites of the National library of France[1]; Numdam[2] for French journals of mathematics; revues.org, Persee.fr[3] which possess journals relating to social science and humanities). There has also been the development of a

1 http://gallica.bnf.fr.
2 http://www.numdam.org/.
3 http://www.persee.fr.

Web TV, Canal-U[4] which today broadcasts 1,153 programs online for tertiary level education, and there are a wide range of diverse teaching resources available.

Different methods of support for the production of digital texts for universities can be taken into consideration:

– since 1998, 150 projects[5] for the development of digital resources and services have been financed, 100 of which are for primary and secondary level education and the remaining 50 are for tertiary level education. The average cost for each project is estimated at 57,000 Euros;

– three invitations to bid (Digital Campuses[6]) were launched for the development of e-learning on a national scale in 2000, 2001 and 2003;

– several initiatives which have been mentioned earlier for French scientific journals;

– it must also be highlighted that the funding provided by the University Agency for the French speaking world has allowed for the creation of an online French library with approximately 3,000 resources available[7].

The different funding provided by the state (French Ministry of Education and Research, French Ministry of Culture) has been used in the following ways:

– each contract which is funded for a period of four years must have a digital resource section included somewhere within the contract;

– the launch of targeted invitations to bid;

– funding applications (aid with the production of resources);

– the attribution of resources distributed by the central administration, for example for the development of thesis services online which is associated with a recently defined regulatory framework[8].

A pioneering phase (1998-2003) supported the diverse and somewhat uncoordinated production method whose investment and continuity did not seem clear initially.

The recent trend, however, has marked the search towards mutualization and a more optimized form of management. The Center for Resources and Information on

---

4 http://www.canal-u.education.fr/.
5 http://www.educnet.education.fr/res/production.htm.
6 http://www.educnet.education.fr/superieur/campus.htm.
7 http://www.infotheque.info.
8 http://www.sup.adc.education.fr/bib/.

Multimedia for Higher Education[9] now has the job of producing and broadcasting these resources (lecture notes, online conferences, picture libraries, etc.) at a national level. Its catalog integrates both types of material, i.e. public material which can be accessed free of charge and resources which have to be paid for, such as those provided by the National Center for Distance Education. This central access point should give more insight into the diverse methods of production which are used. In the same way, productions which are financed by the university agency for the French-speaking world are also organized in an online library as previously mentioned.

The recent creation of virtual universities which specialize in a specific subject is another initiative to increase the production of online resources financed by public funding[10]. The idea here is to structure the provision of e-learning within France and to share the materials produced in this field. One specialized virtual university is currently operational: the French language virtual university[12]; however, the university does not possess a lot of published material.

However, the objective of the mutualization of material remains uncertain. If the mutualization of certain practices is possible in school teaching, it seems less common and more difficult to achieve in higher level education due to the autonomy of the universities and the teachers with regards to their course content since they do not have to follow any pre-defined program so rigorously. The production and distribution of manuals, supporting course material and journals by the university converge towards the idea of a reappropriation of the publishing business. This is a recent trend and it is still not possible to effectively measure the benefits and the quality of the services.

The following issues need to be taken into consideration: the continuation of subsidies, the quality of published material, the integration of the publishing economy within the economy of research and higher education, and the promotion of teachers relative to the amount of teaching materials that they produce. Another option is to pay the teachers for the production of such material (DT-SDTICE 02).

11.1.3. *Published material within a market economy*

Material within a market economy logic covers material which is published by private, French or foreign publishing houses such as journals, monographs, manuals, reference books, bibliographic or factual databases, etc. This type of material is

---

9 http://www.cerimes.education.fr.
10 http://www.educnet.education.fr/superieur/unt.htm.
12 http://www.umvf.prd.fr/.

generally published with the help of financial backing. A large part of the salary of editors and authors comes from the access rights which are paid by the universities or, more rarely, by the final users themselves (students, teachers or researchers).

In fact, very few exhaustive sources exist which show the set of published material used by universities. There is certainly no source which shows the exact conditions of how the access rights are purchased. Nevertheless, by reducing the information made available from the French Ministry of National Education it is possible to have an accurate view of how it works. The French Ministry of National Education carries out a survey each year (ERE) on the electronic resources available at higher education institutions [ERE 03]. This is done through the Couperin consortium which negotiates access conditions to published material on behalf of the universities and through the French Publishers' Association, the latter having carried out a survey related to this subject in 2002 [MIN 02]. With this in mind, two major, intrinsically linked, phenomena clearly stand out: the extreme lack of French material available online and the increasing concentration of resources allocated to the acquisition of digital resources for material in English.

The survey which was led by the French Publisher's Association showed, for example, the lack of online French material and, more particularly, the patterns that have occurred in the following situations:

– the structure of French publishing as far as universities are concerned. This is characterized by a strong atomization and by the presence of a large number of small and medium-sized publishing houses with reduced research and development capabilities. After the Internet boom, the majority of these publishing houses reduced the size of their teams devoted to electronic development, which today makes some people question the existence of expertise within the domain of publishing;

– the limited size of the market, which is a result of the slightly reduced size of the French-speaking area, and the budget constraints of libraries (who have less resources than libraries in English-speaking countries and in countries within Northern Europe), as well as the lack of students purchasing published resources (whether in paper form or in electronic format). Without public help the profitability of the development of electronic products destined for universities seems highly uncertain, which can only prompt economists to err on the side of caution, or indeed to rely on well-known, less risky economic models;

– the French Government has not been able to determine the conditions which are desirable for the development of French material online, neither have they been able to define the support mechanisms necessary for the creation of such material. Except for the French Ministry of Culture, the French Government, and in particular the French Ministry for National Education and Research, seem to have favored the

process of disintermediation. By adopting the principles of disintermediation, the French Ministry of National Education and Research hopes to benefit from the process of digitization by reinstating certain publications within universities that may or may not have been used in their original format, i.e. by digitizing certain paper versions of publications that may or may not have been used before.

In any case, it is important to note that digital material in French which is published online is limited to only two types of services:

– services which are mainly designed for the use of professionals working in universities and in the scientific community, and in particular those people working in the legal and medical fields or in the press;

– services which have been developed by subsidiaries of global groups (Elsevier, Cinven) which are improving the online material available in French-speaking countries by following the example of what is happening in other countries.

Beyond this, the initiatives which are available are rather limited. For example, the availability of the scholarly works of Honoré Champion or the database of journals known as e-Montaigne do not seem to have affected the way in which users access certain information.

On the other hand, it is not an exaggeration to say that it is actually the increasing availability of English language material online which has changed the user habits of those people using French university libraries. As far as journals are concerned, where 98% of the university library budget for digital material is allocated to the purchase of English language resources, the Internet acts as a powerful media source for the broadcasting of international publications to the detriment of local publications, even though the international publications may not be of better quality.

11.1.4. *Published material and a common economy*

The third type of material that we have defined as being part of a common economy is not easy to measure. The main people involved here are the teacher and researcher who, because of a certain professional and political ethic, share their work without asking for any payment. The only thing that they ask for is that the publishers respect the author's ownership of the work. Licenses which indicate the terms of use can also accompany the works that have been made available online. These terms are often based on the Creative Commons license[13].

---

13. http://creativecommons.org/.

These resources are rather dispersed as they are broadcast on the personal sites of the authors who have written them, or on the sites of university departments or associations, i.e. they are broadcast on sites belonging to many individuals or on the sites of various institutions.

Libraries have tried to carry out a census of who uses what resources. A body such as Thot-Cursus[14] links a network of French-speaking correspondents who regularly signal the arrival of new teaching resources. The role of mediators, and in particular libraries and centers of documentation, is extremely important in making this material available online and giving it the attention that it deserves.

As far as the teaching material is concerned, the production of commonly used material that has not been paid for seems to be uneven, and depends on the specific department within the university. For example, this type of production of works is important in the biomedical sector in terms of making classes, books and other different teaching material available online; a census was carried out by the Interuniversity Library of Medicine in Paris[15], by the center of documentation at the University Hospital of Rouen[16] and the multimedia library of the Pasteur Institute[17]. However, the initiative led by the Center for Direct Scientific Communication concerning the sharing of lectures and classes in physics[18] is still not fully developed and only has 26 lessons available on its site. This center was recommended by the scientific community because it hosts the international database of pre-published resources.

The use of such material (explained in the previous section) in teaching tends to be used more at primary and secondary level than in higher education and studies carried out in this field have shown this trend. This trend is marked by the initiatives that can be found on academic websites such as Educasource, Educlic and the sites of associations such as the Pedagogic Café, Webletters, The Clionautes (French association of history and geography teachers created in 1998), Sesamath (an association that aims at giving mathematical and educational documents and software to schools and colleges free of charge), etc.

There are lots of reasons for the imbalance in the use of resources (for which their authors have not been paid) within universities. We give the following hypotheses below in order to explain the reasons behind this:

---

14 Thot/Cursus, http://thot.cursus.edu.
15 http://www.bium.univ-paris5.fr/medecine/debut.htm.
16 http://www.chu-rouen.fr/documed/docum.html.
17 http://www.pasteur.fr/infosci/biblio/textes.html.
18 http://cel.ccsd.cnrs.fr.

– the autonomous habits of the teachers and researchers where there are generally no official teaching programs to follow;

– the authors of the books want their work to be accepted by a recognized publisher in order to improve career prospects for the author;

– the financial interest of the authors as regards the sale of commercially published works.

In terms of the material related to the results of scientific research, the production of shared French material uses the same techniques as those identified at an international level: broadcasting of the material on the personal pages of the researchers, on the websites of laboratories, and also in other more organized forms such as open archives [GAL 03] where the researchers can place their publications. Numerous institutions have created open archives of their publications. The Center for Direct Scientific Communication (CCSD/CNRS) offers its services to the entire scientific community; however, it is only responsible for three French archives[19].

This publication of material remains heterogenous and often too widespread. After the innovation phase, the risk of failure becomes larger and the need for a recognized publisher to accept the book becomes increasingly important.

11.2. Published digital resources and distance teaching devices: an even weaker synergy

Linking the digital resources which are available to the teaching profession seems legitimate when thinking of the end users of such resources (the students), especially when a particular teaching body uses the Internet as the main method for communicating with and teaching its students. The significant development of distance teaching over recent years has led to questions being asked about the effective integration of these various published resources in distance teaching, also known as virtual campuses; there are slight differences in the different terms which are used [GLI 02].

The devices used in distance teaching integrate several functions and variables depending on whether the activities take place on campus or are available by remote access. The different variables include: the personalized monitoring of the learner, the collaborative work tools available between learners, and the resources made available by the network. These resources generally become more varied with information relating to the co-ordination of the teaching (timetables and the different grades of the learners).

---

19. http://www.ccsd.cnrs.fr.

The research[20] that we are currently leading along with a sample of online campuses essentially shows the presence of teaching and administrative resources and an absence of documentary and published resources in these virtual campuses. There are only some links to the campus library catalogs, which is quite paradoxical because we are aware of the different journals that were made available online by the libraries that purchased the access rights for their university. The link between the digital library and distance teaching still obviously remains a problem in France; there is insufficient communication and synergy between librarians and teachers. Is this simply the reflection of current teaching practices in France, characterized by a weakness of recommended reading for students [MIN 02]? The comparison with other countries remains to be investigated in order to find out whether this is a phenomenon that is specific to France.

Another observation that has been made deals with the current development in the quality of teaching resources available. These resources are considered as published resources insofar as they are generally validated by councils within the teaching institutions.

The survey carried out in 2003 by Ernst & Young in partnership with IPSOS [ERN 03] which evaluated French online campuses gave the following results:

– 74% of the teaching resources were produced within the university;

– 20% of the resources used for teaching were also used for another purpose;

– only 6% of the resources were purchased from outside the university. Within this 6% of resources purchased from outside the university, only 45% of the total came from publishers, the remainder of the resources came from other French and foreign establishments.

If we convert these figures into the share of the budget as regards the total cost of online resources in France, 39% of the budget is allocated to producing documents within the university and 2% is allocated to the purchase of documents from outside the university.

This observation partly reflects the new publishing framework which is becoming more and more apparent for teaching resources: the local production of texts is dominant, the sharing of documents amongst teaching establishments remains weak, and collaboration with private editors is practically non-existent. Is the current reality that it is no longer necessary to depend on an external publisher? Cost assessment, as well as an assessment of the quality of publishing and of the teachers and authors, must be completed before discarding external negotiations

---

20 Béatrice Baczkowski, PhD (Director G. Chartron), Doctorate School of IT and Computer Science for Society, July 2004.

which up until now have led to the monitoring and adaptation of teaching material for different types of students. New scenarios which develop partnerships between universities and publishers are not really being considered. This means that it would no longer be a question of publishing books or journals that are purchased less and less by libraries, but rather a question of developing material that is produced by new economic methods.

11.3. The evolution of activities for libraries: future priorities?

11.3.1. *Evolution of activities*

Inevitably the increasing availability of digital resources has forced librarians to rethink their roles [CHA 02] as well as the services they provide relating to the new needs of remote users. Strategic guides aimed at professionals have been written to accompany these organizational changes [ACR 00].

Different studies carried out have identified a common core of significant services and standards relating to the expectations of online users [MUT 03]:

– access to online material;

– services which provide documents on demand for documents which are not available;

– different support services: online guides, questions and answer services by email and telephone.

All of the activities which affect librarians in the development of their jobs can be summarized in Table 11.1. The back office shows the fundamental activities which are centered on the make-up of the stocks of resources, and for which the number of different services available to the users (which can be seen in the front-office) have declined.

Back office (life cycle of documents)	Front office (services → users)
Purchasing policy Purchasing, negotiation of licensed resources Publication of local resources Archiving and referencing of documents	Technical access tools (interface, search engines)
	Organization of material (index, access keys)
	Broadcasting the material (alerts, newsletters with news of the latest events)
	User support (guides, information literacy, question and answer services)

Table 11.1. *Back office/front office activities*

In terms of digitization, the back office is characterized by a hybridization of resources which must be taken into consideration nowadays. Jean-Michel Salaün distinguishes between individual resources, licensed resources, specially created resources and free resources [SAL 01]. Depending on the scientific domain, the evolution and allocation of these resources is different as pointed out by Cherifa Boukacem [BOU 04]. The last few years has seen the development of renewed activities of libraries in the back office, primarily for:

– the negotiation of access licenses and the appropriation of new management modalities;

– the development of publishing (as well as intellectual and technical) activities to highlight local resources, and the digitization of ancient collections of works, student theses and dissertations online, as well as conferences held at the university. However, this development is uneven depending on the library and depends on available human and financial means as well as local partnerships.

In France the front office has been particularly marked these past few years with the development of access interfaces, the renewal of location tools which organize the attention economy[21], and the development of certain tools aimed at accompanying the user (such as the creation of class supports[22], training and the emergence of the first online question and answer support[23]). The new modalities for broadcasting the material, however, still need to be developed in many cases: only a few libraries publish a newsletter, organize general broadcasts or consider using automatic broadcasting techniques in order to develop and broadcast their latest news.

11.3.2. *Two future priorities for libraries*

11.3.2.1. *Developing links with pedagogy*

There is currently a lack of collaborative work between teachers and librarians in France. This idea of teachers and librarians working together is essential, even if it is not necessarily new.

Whether the collections are on paper or are in digital format, the assumption is still the same; learning does not have to be limited to the classes given by the

21 For example the SUDOC (University System of Documentation) catalog: http://www.sudoc.abes.fr; as well as all the bookmarks selected by libraries and those of the National Library of France.

22 For example visit the Formist website in order to have a closer look at the effort made to try and share class supports, http://formist.enssib.fr/.

23 Log on to the question and answer support of the Interuniversity Library of Medicine in Paris, http://www.bium.univ-paris5.fr/medecine/refvirt_cad_ang.htm.

teacher. The increased stocks of books and other resources in libraries has made it possible for researchers, analysts and people who like to read to widen their knowledge about a specific subject.

Digitization offers new possibilities in terms of accessing material which, up until now, has been limited by the hard copies of books etc. that are available. The negotiation of access rights must be discussed amongst all partners, but the issue is more of size since it is about enlarging the reading practices and habits of students. Libraries are at the center of this pedagogic issue: they should be involved in the elaboration of all distance teaching plans because nowadays it is a question of creating advances in teaching, as well as creating interconnections between teaching establishments which have been made possible by the dematerialization of the online content. It has not yet been established whether the student would be able to access all of these online resources without the help that is provided by the libraries.

It is also necessary to monitor the development of the autonomy of this information. From a teaching point of view, it is interesting to see if the users of these resources are able to master and use all the information that is available to them, i.e. monitor how distance learners use the sources, instruments and how they process the information. It may also be necessary to provide some type of training for them. This is essential and could lead to the creation of specific teaching modules which the libraries could be responsible for rather than creating simple guides.

On the other hand, the link between the library and teaching process is essential, especially when taking the life cycle of the online material into consideration. How is it possible to list the material, insure its continuity and continually update it within the university context? This issue is more a technical and organizational one, but it is fundamental in ensuring a certain level of continuity and quality of developed services.

11.3.2.2. *Developing partnerships for the creation of stocks of digital material*

The way in which online scientific and teaching material for universities is produced within the universities themselves remains a delicate issue for several reasons.

First of all, publishing is a job which requires a certain level of acquired know-how and it would certainly be unadvisable not to use this know-how in the creation of digital resources. Furthermore, the complete integration of publishing within a teaching institution might exclude other publishers from being able to critique future resources that are to be published. This may also lead to the fact that too many teachers rather than the publishers who have experience in this field become involved in the publishing process. Finally, in relation to the publications released by the French University Press, it can be said that different people have different views on the works that it has produced and this suggests that the French University

Press encounters the same problems for publications that are used both within and outside French universities.

The scenarios would therefore have to be diverse and would have to create partnerships with publishers in order to produce an appropriate amount of online material in the non-English-speaking world. The provision of this online material, which is based on an economic model respecting the interests of each party, is an approach that could be investigated in the future.

The Manum project [MIN 03] was developed following such a logic. It was a project linking publishers, librarians and teachers to create a digital library with documents which would be used as teaching tools for students of Bachelor's and Master's degree programs in humanities and social sciences, as well as for their teachers. The project consisted of introducing a one-stop shop which allowed for the digitization of publishers' documents on request, as well as negotiating the associated access rights. The project has not received any real attention from the French Government. It is just a question of waiting to see what partnerships will develop over the years to come.

11.4. Bibliography

[ACR 00] ACRL, Guidelines for Distance Learning Library Services, http://www.ala.org/ala/acrl/acrlstandards/guidelinesdistance.htm, 2000.

[BOU 04] BOUKACEM C., *Du prêt entre bibliothèques à la bibliothèque virtuelle: une évolution contrastée des bibliothèques universitaires françaises selon la discipline*, Doctoral thesis, University of Lyon 2, 2004.

[CHA 02] CHARTRON G., *Les chercheurs et la documentation numérique, nouveaux services et usages*, Publications of the Cercle de la Librairie, Library Collection, 2002.

[DEC 02] DECEUNINCK J., "Campus numériques, établissements virtuels, ressources en ligne : la carte de l'offre et le parcours de l'usager", *2001 Bugs, Globalism and Pluralism*, Montreal, 24th-27th April 2002, GRICIS, 2002.

[DTS 02] DT-SDTICE, Legal framework – Online campus DT-SDTICE, April 2002, http://www.educnet.education.fr/chrgt/juricampus.pdf.

[ERE 03] ERE, Surveys carried out in 2003 on the electronic resources purchased in the Common service of documentation and in Information Systems and Management Control and also in higher education libraries, http://www.sup.adc.education.fr/bib/Acti/Electro/accueil.htm.

[ERN 03] ERNST & YOUNG, IPSOS, Evaluation of online campuses, http://www.educnet.education.fr/superieur/CN-evaluation.htm, 2003, and the Power Point presentation in particular: http://www.educnet.education.fr/chrgt/CN-docipsos.ppt.

[GAL 03] GALLEZOT G., ROSSI C., CHARTRON G., NOYER J.-M., "Conception d'une archive ouverte en SIC: le sens de la technique", *H2PTM'03, Créer du sens à l'ère numérique, Hypertextes, Hypermedia*, University of Paris 8, 24th–26th September 2003, http://archivesic.ccsd.cnrs.fr/sic_00000943.html.

[GLI 02] GLICKMAN V., *Des cours par correspondance au e-learning*, PUF, 2002.

[MIN 02] MINON M., with the collaboration of A. LANGLOIS-MEURINNE and E. NEU, *Publication universitaire et perspective du numérique, study carried out for the Syndicat National de Publication*, http://www.sne.fr/numerique/, 2002.

[MIN 03] MINON M., Manum, Experimental project for digital libraries in human and social sciences, http://manum.enssib.fr/, 2003.

[MUT 03] MUTINTA MAYO L., STERN-CAHOY E., "Meeting the needs of remote library users", *Library Management*, vol. 24, no. 6/7, p. 281-290, http://www.emeraldinsight.com/0143-5124.htm, 2003.

[SAL 01] SALAÜN J.-M., "Publication, édition, accès, quel modèle économique pour les médiateurs ?", *Congrès IDT 2001*, Paris, 2001.

[SCE 04] SCEREN-CNDP, "Les nouveaux espaces numériques d'éducation", *Files on Education Engineering*, no. 46, March 2004.

[SCH 97] SCHWUCHOW W., "L'information : bien public ou privé?" in A. Mayère (ed.), *La Société Informationnelle*, L'Harmattan, p.113-123, 1997.

Chapter 12

The Revel@Nice Project:
the Creation and Prospects of a Pioneering Site
of Online Periodicals and Journals

The website Revel@Nice is a website that belongs to the University of Nice and is made up of online journals and periodicals. The website was created from the belief that it would be useful for the university to have a website on which the university's research teams could display their online journals. The website was also created to provide solutions to specific problems that could only be answered on this site. If the introduction of this site was a technological achievement, the project itself was more of a functional and organizational achievement.

This chapter presents the genesis of the project, its creation and its current prospects in terms of longevity and sustainability.

12.1. The project

12.1.1. *Purpose of the site*

The need for such a project is based on the observation of four facts:

– the need for our university to have a system which clearly displays its research, starting with the production of its documentation;

Chapter written by Michel ROLAND.

– the need for the research community to adopt an active attitude in relation to the larger scientific publishers whose monopolistic position is even threatening the distribution conditions of scientific documents;

– the necessary complimentarity which is needed between universities and research organizations in the field of research. This complimentarity is present in a variety of forms depending on the subject and on the segments of the documentation chain. This means that at the local level, production (there is a need for training and support within research teams) is more relevant than distribution (where disciplinary logic is more relevant);

– the proximity between the domain of online publishing and that of the acquisition of online material, of which the research department (university library) is in charge. This proximity can be observed at the strategic level (scientific publishers) as well as at the technical level (common technical data and relevant standards as well as the signaling of online documentation and the production of such material).

However, the project also responds to a particular economic situation, that of the publishing crisis which the paper versions of journals are currently experiencing at the Faculty of Humanities at Nice. At the University of Nice a publishing association has been in existence since 1967. It is a voluntary association known as the Publishing Association of the Faculty of Humanities of Nice (l'Association des Publications de la Faculté de Lettres de Nice or APFLN). This publishing association has published almost 200 titles since its creation back in 1967. The titles which have been published primarily cover a broad range of academic works such as monographs, theses, seminars, deeds and festschrifts (of which there are ten journals). All of these titles have been distributed since 1996 by the Inter-institutional Center for the Distribution of Publications in the Humanities (Centre interinstitutionnel pour la diffusion de publications en sciences humaines or CID), which is a specialized service by the Maison des Sciences de l'Homme in Paris (the Maison des Sciences de l'Homme is a French foundation which aims to promote the study of the humanities). In the absence of a University Press, the APFLN acted as the interface between the research centers and the publisher. In 2001 a step was taken within the Faculty of Arts and Humanities at the University of Nice to create the Nice University Press (Presse Universitaire de Nice Sophia Antipolis or PUNSA). The project did not last long as it was felt that there was no need for it on the other campuses of the university. In 2002, the CID stopped distributing for the APFLN for both organizational and economic reasons. Therefore, the suggestion of a site for electronic journals played a part in the serious crisis regarding the paper versions of these journals.

12.1.2. *History*

The project was subject to a double presentation; by the director of the research department (Louis Klee) and the author of this chapter (Michel Roland) who was the person in charge of online documentation during the two consecutive sessions of the Science Council of the Sophia Antipolis University in Nice on 25th October 2002 and 22nd November 2002.

On 25th October 2002 the proposed project was defined in its general principles and placed in its technological, functional and organizational context.

The principles of the project were:

– on the one hand, to integrate the project within the information system of the university, and in particular within the framework of an institutional warehouse, followed by the creation of a website of online theses and dissertations, a website of archives of pre- and post-publications and of possible online collections of work;

– on the other hand, to implement the use of known formats within the systems (XML and OAI).

On 22nd November 2002 the university science council provided funding of €22,000 for quality research bonus publications.

The principles of the project were divided into three different parts, each different part acting as a constraint to the development of the program:

– offering authors at least the same conditions as traditional methods offer (such as paper) in terms of visibility and validation;

– designing and developing the "electronic journals" site as a building block within the information system of the University of Nice;

– developing the correct formats which ensure the visibility and interoperability of the publications.

In order to overcome the problems associated with the constraints mentioned above, the designers of the project had hoped to use the expertise of a service company. In order for a service company to help the project designers there would have to be a review of all the resources that could be published online.

The project that was launched encountered two main problems:

– a lack of adaptability and resourcefulness of the university (review of resources);

– the preliminary meetings with the service companies highlighted the inadequacy of the budget.

Several measures were taken to try and overcome the financial difficulties:

– searching for partners to join the project: contact was made with the University Franche-Comté (Besançon) which had introduced an evaluation and signaling portal from its research. It was believed that the two projects seemed to complement one another, however, they were both too detailed in their own field to be linked together. The idea of joining the two projects together only came about after they had been created. A call for mutualizing the two projects was launched through Couperin's list (Couperin's list is a list of online journals available at the University of Rennes 1). The idea was to ignore the temporal nature of universities; however, no active partnership was created in this way;

– another measure arose with the invitation to bid for the creation of a retrospective portal of online journals within the humanities and social sciences. This invitation to bid was launched by the subdivision of the libraries at the French Ministry of National Education. This project which has been known as Persee ever since it was created (http://www.persee.fr) has offered to display online all of the archives of the seven most prestigious French journals in the area of humanities. Working on this project re-ignited the ambition for the project at the University of Nice to work. The objective of the national portal largely covered the objective of the local portal, however, the main difference is found at the national level as regards the acquisition of material: a retrospective portal implies acquisition by digitization whereas a common portal assumes a publishing interface to be used by the authors of articles. It was therefore possible to re-develop and re-design the site of the University of Nice as an additional development to those carried out in the framework of the national project. This choice naturally led to a lack of ambition as regards the Nice project, in particular as far as time was concerned: the creation of the Nice site was delayed for at least a year. Nevertheless, in May 2003, the University of Nice teamed up with the University of Lyon 2 to work on the creation of the Revel@Nice project. The University of Lyon 2 was looking for a partner to work on its project at the time.

However, in June 2003, the Science Council at the University of Nice decided to give the remainder of its subsidies for publishing to the Revel@Nice project. These subsidies had not been distributed previously. They decided to dedicate the money to the project because of the crisis affecting the APFLN regarding the distribution of their publications. The sum of money that was allocated to the project was €57,500, which was almost two and a half times more than what was previously donated to the project. All of a sudden the prospects of the creation of a local site once again became a reality, however, with considerable time constraints (the subsidies had to be used by the end of 2004).

12.2. Creation

On 12th June 2003 an invitation to bid was published at the university. The invitation to bid was aimed at research teams and came from science departments 6 and 7. The science departments asked the candidate teams to apply before 15th September 2003.

The text of this invitation to bid detailed the respective obligations of the various parties involved in the project: the science council, the team building the website, as well as the research and publishing teams:

> The science council is offering candidate teams the opportunity to create an online publication of their journals. The science council will maintain the site and support the teams for all operations linked to the online publication.

> The site is part of the documentation information system of the University of Nice which is part of a four year contract. The site will be created according to pre-established standards, in particular the archives will be created according to recognized formats which are interoperable with national projects (such as the portal for humanities and social sciences) as well as existing international portals (Erudit and OAI, etc.).

> The user interface will respect the identity of the journal and its publishing committee.

> Storage and archive techniques will ensure complete compatibility with a possible paper edition (which is not the subject of the current invitation to bid) and will make the creation of a possible paper edition easier by the addition of a sequence of specific processes.

> An ergonomic interface for the use of authors and for validation procedures (by the publishing committee) will also be created. In December 2004 the teams will put their final editions online on the site which will include all of the essential functions. The editions which are released after the site is up and running will be added to the website so that users of the website can also have access to the latest versions of the online editions.

> Within a two year deadline a digitization plan will ensure that journal archives (old editions) are made available online.

> The candidate teams will commit to whole-heartedly working with the team responsible for the creation of the site, which is elected by the science council, and will also commit to respecting working procedures.

> The candidate teams will appoint one person within their team to represent the research team within the team which is responsible for the creation of the site.

Questions in terms of rights and conditions of access to the articles will be decided in collaboration with the publishing team and with the team responsible for the creation of the site.

After taking the invitation to bid into consideration, some important issues still remained, such as:

– access to the site is free. The possibility of a subscription is not excluded but postponed for technical reasons;

– if a journal is put online, it is imperative that it is the latest edition of the journal;

– even if teams are asked to appoint a person in charge of the publication, it is not exactly clear what the differences in skills actually are between the publishing team and the team responsible for the creation of the site.

10 teams (and their respective journals) responded to the invitation to bid:

– Noesis (Philosophy);

– Corpus (Linguistics);

– Loxias (Humanities);

– cahiers de l'URMIS, which is an online journal published by the research unit of migration and society (Sociology);

– socio-anthropology (Anthropology);

– Journal of industrial economics;

– cahiers de la Méditerranée, which is an online journal published by the Center of the Contemporary and Modern Mediterranean (History);

– cahiers de narratologie, which is an online journal of the theory and analysis of cultural, literary and artistic productions (Humanities);

– Cycnos (English studies);

– Journal of Alpine history and archaeology.

The journals are available in almost all types of possible format: existing journals published in paper format within the framework of the APFLN, journals in paper format published by traditional commercial publishers, a journal which already exists online but in raw html and a new journal.

12.2.1. *Timetable for the creation of the site*

The basic period of the subsidies meant that the site had to be created before the end of 2003, which was a deadline of only six and a half months from the moment the science council donated its funds to the project. Given the deadlines for recruitment and the particularities of the university year, the development and the definitive creation of the site would have to be carried out in a four month period, from September to December 2003. We gave ourselves the target of having an operational site with journals online on 15th December 2003 and with an official launch of the site on 1st January 2004.

The details of the timetable were defined as follows:

– July–September 2003: precise definition of the project and of the principles of the website's structure. Pre-selection of temporary employees (profiles and adverts on biblio.fr and adbu-forum etc.);

– September 2003: recruitment of the team of temporary employees;

– 1st September–25th November 2003: definition of procedures, formats and the website's structure;

– 1st October–15th November 2003: writing of the acquisition interface;

– 15th October–15th December 2003: writing of the user interface (portal);

– 1st November–1st December 2003: training authors how to use the tools;

– between 1st December and 15th December 2003: the portal goes online.

12.2.2. *Human resources and project management*

The choice was taken to create the site internally using specially recruited temporary employees for the project:

– system administrator: in charge of the infrastructure and structure of the site (Stéphane Chauvin);

– system developer: in charge of the creation of specific documents (interfaces) (Denis Méline);

– project engineer: in charge of contacts with the publishing teams, as well as the daily running of the project and the training of authors and publishing committees on how to use the tools associated with the site (Jean-Vincent Holeindre).

Staff from the research department were responsible for the supervision of the work:

– Michel Roland, a librarian in charge of online documentation with the research department, was in charge of the project. For doing this the Presidency of the university made him the person responsible of the project's "electronic publications research";

– Ghislain Chave, head of the IT service within the research department was in charge of the technical management of the project.

Sophie Rapetti, assistant to Michel Roland, was included in the team to monitor the logistics and co-ordination of the project.

A guidance committee was also created and was responsible for discussing, validating and controlling the execution of the project during each of the important steps of the project. The guidance committee was made up of seven people, each chosen because of their expertise in the field of online publication as well as their expertise in Information and Communication Technologies for Education (Technologies de l'information et de la Communication pour l'Éducation or TICE) or in the relevant scientific domains. The heads of the Faculties of Law and Humanities, as well as a representative from the science council, were also part of the guidance committee.

Listed below are three points which take the organization of the project into consideration:

– the involvement of staff from the research department as well as the guidance committee and the management of the project portrayed the project as relying on the university's research department. On the other hand, as far as the governance of the organization of the project was concerned, it could be said the project was linked more to the university's science council;

– exchanges with the heads of similar projects (in particular with Odile Arthur for Pelleas in Marne-la-Vallée) made us aware of the risk of under-estimating the non-technological aspects of the project, in particular the training of the users and producers (the authors) for which one of the three non-technical members of the team had to be recruited;

– the entire guidance committee was made up of people from Nice which was unfortunate, but because of time constraints it was impossible to have it any other way.

12.2.3. *Precisions and modifications*

Throughout the project's development we were forced to specify and modify some of the objectives that had been initially established:

– the humanities and social sciences' portal which was outlined by the University of Lyon 2 in participation with the University of Nice was not launched in the end. Certain technological choices had to be made for the creation of this portal; these choices were similar to those made for the national project and therefore this more localized project could not be launched. This was the case for the reference document type definition (DFD): the DFD "Erudit" was part of the requirements for the national portal;

– exchanges with publishing teams enabled us to have a large selection of published texts, i.e. we had access to the same texts in both paper and electronic formats; these exchanges with the publishing teams made it possible to make editions of journals online, and to publish a summary or analysis of the latest edition of the journal. The fact that we had access to the same texts in both their paper and electronic format was an important factor in ensuring the teams' support for the project by providing a certain level of complimentarity between the online version of the journal and the possible commercial distribution of the paper version (some people believed that we were favoring the use of the online version of the journals over the paper version);

– the involvement of the graphic/web designer was underestimated at the beginning, and as a result the budget devoted to the work of the graphic/web designer was doubled. On the contrary, the budget for hardware was three times less than what had been initially estimated;

– finally, after the definition of the interfaces and functionalities of the site it was decided that the open source software Lodel would be used for the production interface. Lodel was developed for the website revues.org despite the differences in opinion between the various technological parties involved. Whilst local developments are based on Java technology, Lodel is based on PHP technology; in addition the integration of XML within websites was less advanced at the time. The problem that we faced was as follows: our developments would undoubtedly be finished in time for the launch of the site in the middle of December 2003. However, in order for the publishing teams to be trained on how to use the tool and how it worked, followed by the process of putting the site online, the publishing teams needed to have access to the tool from the beginning of November 2003 onwards.

These precisions and modifications were validated by the guidance committee on 28th October 2003.

12.2.4. *Launch of the site and performance analysis*

On 15th December 2003 9 out of the 10 journals were online and the site was launched, its official launch occurred on 1st January 2004 as planned. The review of the project's performance was presented to the science council of the University of Nice on 30th January 2004. The project was completed within the planned timescale, following the established requirements and their modifications mentioned above. The project was also completed with a budget that was much less than what was allocated to it, which made it possible to extend the contracts of the employees over the first four months of 2004. Taking into account the fact that the Revel@Nice project is the first website of its kind (an institutional platform for the displaying of electronic journals online) and the experimental nature of the project, we can say that the "trial" has been a success. However, the site needs to be transformed so it can be used in the long-term.

Figure 12.1. *Homepage of the Revel@Nice electronic journal website*

12.3. Sustainability and longevity

The issue of sustainability can be divided into three different sections (March 2004):

– editorial and publishing sustainability;

– technological sustainability;

– institutional sustainability.

12.3.1. *Publishing sustainability*

State of the art tools are used in the production of content online. These tools are also available to the publishing teams of the journals. These tools also give the maximum amount of autonomy possible to the journals. However, the competency levels as regards the use of information technology by the teams, as well as the suitability level of the publishing tools used by the teams, are very different. If the teams do not have all the human resources necessary for the publication of their journals, it is then desirable to have some sort of publishing aid within the site. This becomes more important in terms of monitoring the arrival of new journals appearing on the website, in other words the publishing aid enables the teams to determine whether the new journals can be published on the website or not (see guidance committee).

We can see the need for a full-time position in this field. Such a position falls within the competences of the publishing press which already exists at the University of Nice.

12.3.2. *Technological sustainability*

The computing tools and interface which were developed for the website do not constitute a complete and closed system. The change from PHP technology (Lodel) to Java technology makes it easier to integrate XML within the site, which is currently being worked upon.

We are also committed to developing and improving the current interface and ergonomics of the site (by taking into account users' opinions, online surveys and user forum discussions). In this respect, what is particularly important is the work we are involved in, which is aimed at improving the publication of articles in PDF format (poetry layout, mathematical formulae, etc.). This work is carried out with the University of Nice's participation in the national project (portal for journals within the humanities and social sciences) and is led by the University of Lyon 2. In

the longer term the need for technological developments for the website itself should become less and less important, but the site will still need continual maintenance, will still need to evolve and will still need to provide support to its users. On the other hand, developments for the site should be replaced with the organization and installation of an institutional open archive, which has been requested by users.

As far as technological sustainability is concerned we can see the need for a part-time position before work commences on the institutional archive, and after this has been created a full-time post will be required.

12.3.3. *Institutional sustainability*

The guidance committee which monitored and validated the development of the project should be replaced by two distinct committees: a monitoring committee which would be in charge of validating the technological evolutions of the site and a publishing committee which would be in charge of the scientific quality of the journals made available online as a complete set (the appropriate publishing and editorial teams are capable of producing the scientific content of each of the journals). The publishing committee will also examine and validate the applications from new journals that want to be published on the site.

The composition of the monitoring committee can be the same as or similar to the make-up of the guidance committee. The composition of the publishing committee is still to be drawn up, taking inspiration from the make-up of publishing committees of similar sites such as revues.org or Erudit.

However, the website and its associated committees must clearly be part of the university. The electronic journal website must be considered within the context of the information system of the University Sophia Antipolis in Nice, and more precisely within the framework of an institutional open archive. The Revel@Nice project is the leading building block of a publication tool used for the results of research carried out at the university. The growth of this institutional archive and its integration within the computing service of the university assumes a continuous commitment to research publications.

The Revel@Nice project was created with the administrative support and infrastructure of the university's research department. The project benefited from the supervision of the research department's computer specialist. The issue of integrating this responsibility within the research department as far as the traditional roles of the research department are concerned (with the exception of the theses service) means that this added new responsibility will affect the acquisition and distribution of documentation and not the production of such documentation, both

within and outside the research department. The solution could be to join this new service to the presidency and scientific council of the University of Nice with the long-term goal of completely integrating the service within the research department.

In order to assure this sustainability, in 2004 a finance plan was drawn up and validated. The finance plan combined the central structures of the university and the different faculties, up to the investment of their laboratories on the site.

12.4. Post-scriptum: today

Between April 2004 and December 2004 the Revel@Nice project was working in slow motion. Only a part of the finance plan was received. Of the project's three engineers, only one computer specialist will remain from March 2005 onwards, who is largely devoted to working on the national project (portal for journals within humanities and social sciences/Persée). The absence of publishing and editorial support has largely left the editing teams somewhat isolated.

12.4.1. *Visibility*

At the same time, statistics showing the number of hits on the site (approximately 250,000 visits and 750,000 pages read in the space of a year for a total of 386 articles in nine journals) confirms the success of the site and the strong impact that it has had on the articles which are available online. There are links to the site on numerous specialist sites, and the site is recommended for the journals that it possesses. An agreement was reached with revues.org on the basis that revues.org would non-exclusively display the contents of the journals of the University of Nice and would do so free of charge. This meant that the journals published by the University of Nice could now be accessed from a major reference site. The compatibility of these sites with OAI meant that the journals published by the University of Nice could now be taken into consideration by harvesters (see OAIster: http://oaister.umdl.umich.edu).

12.4.2. *Versioning*

As soon as the site was launched online a questionnaire was integrated in the interface to gather the opinions and reactions of users in order to improve the interface and to prepare a new version of the site. In the spring of 2004 Denis Méline and Sophie Rapetti organized a series of interviews and tests following a standard method. This was carried out in collaboration with researchers from the Maison des Sciences de l'Homme. The process which was defined was then used within the framework of the national project (Persée) and the results of the tests

were integrated along with the development of the functionalities of the national project.

12.4.3. *Current prospects and perspectives*

Today, the team has been expanded with the hiring of Khatir Aib, a computer specialist in October 2004, and Marie-Luce Rauzy, an editing secretary, on 1st December 2004. Marie-Luce was hired thanks to the Regional Unit for Training and Promotion of Technical and Scientific Information in the Provence-Alpes-Côte d'Azur region (URFIST PACA). Developments to the site have recommenced and the integration of new journals (international law and stylistics) was scheduled in January 2005.

Revel@Nice is also part of the project launched in November 2004 to organize a platform of French-speaking networks of scientific journals. Other partners in the project were Erudit and Persée (the project leaders), revues.org, the CNES, the portal of the CNRS for reference journals in the humanities and social sciences, the Franco-Belgian project known as Cairn (University of Louvain) and the Belgian project Peps. The joint work between Persée and the Québécois team Erudit on the DTD/Erudit system allowed the groups working on this larger project to ensure that our journals were made available online in the best technical conditions possible. The joint work between the two groups also enables us to locate where we are as far as non-commercial platforms for online electronic journals are concerned.

Nevertheless, in order to have a certain level of sustainability the Revel platform has gone through some changes, notably the sharing of the system with other institutions. Revel does not only answer the technical and resource needs of the University of Nice, and because of this fact our aim is now to share our experience and the creation of our site with other universities. Sharing the platform has allowed us to free up more important resources. Sharing the system can be undertaken in two levels: first of all on a regional level which means that not too many different groups of people are involved in the creation of the system both on a technical and editorial level. The other method lies at the national level where the sharing of the platform would be limited to the system's tools and associated developments. Contact with other institutions is currently underway so that we can move forward in such a way.

However, sharing the system demands a clear awareness of the specifics of Revel and of its insertion within the framework of electronic journals.

Although Revel@Nice is an institutional website, it is above all a tool which has been created by one university. The website is used particularly in humanities and social sciences, law and economic sciences and in the natural sciences and medicine.

The last two are slightly more integrated in the global network (predominantly English-speaking) of large international scientific publishers. In terms of the documentation resulting from research, the purpose of the site is more about the production of the documentation rather than its appearance online. During the creation of the website it was noted that the groups produced journals and articles that were relevant to the level of the university students and personnel. In respecting the standards of interoperability, coupled with the development of partnerships, it was therefore possible to combine the production and distribution of journals and articles. This complimentarity, which is necessary from a researcher's point of view, however, does not transform the distribution interface into a simple appendix of the production interface. This idea of combining the production and distribution of journals and articles comes from the integration of the Revel project within the information system of the University of Nice. This integration also makes it possible to display what is happening in the area of science at the university, as well as motivating and stimulating research groups within the humanities and social sciences. The integration of the website within the university's information system also makes it possible to modernize publishing methods stemming from the results of studies carried out by the research groups. This integration is not passive, which can be seen by the exchange of data with other sub-systems within the information service of the university, such as the development of research, teaching resources, etc.

From an internal point of view (the development of the information service within the University of Nice) as well as from an external point of view (sharing of the system and partnerships), the Revel@Nice project is simply the beginning and our problem today is trying to work out how we can advance along the many paths that have opened up before us.

This page intentionally left blank

Chapter 13

Evaluating the Use and Users of Digital Journal Libraries

13.1. Introduction

CIBER have spent the last five years evaluating the use and users of digital journal libraries or publisher platforms, as publishers prefer to refer to them, as part of an extensive virtual scholar research programme. We have done this with a methodology fit for purpose, i.e. digital methodology, which is suitable for evaluating millions of global scholars who themselves make tens of millions of transactions [NIC 05a]. This methodology is generally referred to as transactional log analysis, but more appropriately, as deep log analysis in recognition of the special way CIBER employ log analysis and the fact that the logs are related to user demographic and attitudinal data. What we have done for the purpose of this chapter is to pool the knowledge we have gleaned from these studies to provide robust evidence and establish some universal truths about the nature and behavior of the scholarly users of digital journal libraries – the virtual scholars.

Given the importance of the methodology in delivering robust evidence-based data for literally hundreds of thousands of virtual scholars, we shall first spend a little time explaining the methodology. Log analysis has a number of characteristics to commend itself to researchers of digital libraries:

– *size and reach*: logs provide an enormous reach and there is no need to take a sample (and many digital library studies have been weakened by small and unrepresentative samples). It is possible for a single researcher to evaluate the use of

Chapter written by David NICHOLAS and Paul HUNTINGTON.

hundreds of thousands of users, and they can do this remotely, and without intrusion;

   – *they are a direct and immediately available record* of what people have done: not what they say they might, or would, do; not what they were prompted to say, not what they thought they did. Logs provide a powerful evidence base;

   – *they provide real-time and continuous data*: logs create a digital laboratory environment in which innovation can be introduced and its impacts immediately monitored and measured.

   Deep log methods build upon these attractions and provide for a much more robust analysis of raw transactional server log files and the linkage of these data to user information datasets to provide important triangulation. In the simplest terms the digital usage fingerprint (the logs) that users leave behind them as a result of visiting a digital service are deeply mined and enhanced by relating user information to the fingerprints, thus considerably enhancing the yield of information that can be obtained. With traditional log studies we learn (crudely) about use (hits) rather than users and, as a consequence, are left with generalizations about hundreds of thousands of users, amongst who may be students and Nobel Prize winners, which is, clearly, not very useful information. In the recent past CIBER have, in a study of Blackwell's Synergy, related use data to the user data that can be obtained from the IP address and a subscriber database [NIC 06a]. However, in our most recent study, that of Elsevier's ScienceDirect, usage has been related to user demographic and attitudinal data obtained from online questionnaires [NIC 06b].

   The types of analysis deep log analysis can produce for digital libraries are as follows:

   – *use analysis*: by number of views, number of search sessions conducted, site penetration (number of views made during a session), repeat visits made, time spent online, and type of view (TOC, abstract, full-text);

   – *user analysis*: by age, gender, occupation (student, practitioner) organizational affiliation, heavy/light user, referral link used, type of university (research/teaching), subject/discipline of journal, subject discipline of the user, search approach adopted, geographical location, academic status (student, professor, etc.), whether articles purchased online or not, article productivity and use of additional database functions.

13.2. Digital libraries evaluated

The CIBER virtual scholar research programme has provided evaluations of the following six digital journal libraries (in ascending data order of study, with name, funder and date of the research project in brackets):

– *EmeraldInsight (www.emeraldinsight.com)*, a database of journals from the fields of business, management and library studies *(Digital Journals – Site Licensing, Library Consortia Deals and Journal Use Statistics*, the Ingenta Institute, 2002);

– *Blackwell Synergy (www.blackwell-synergy.com)*, database of learned society journals (*Virtual Scholar Research Programme – Use and Impact of Digital Libraries in Academe*, Blackwell/Emerald, 2003-2004);

– *Institute of Physics Publishing database (http://www.iop.org/EJ)*, a database of more than 70 physics titles published by IoP (*Physics Journals: A Deep Log Analysis of IoPP Journals*, Institute of Physics, 2005-2006);

– *OhioLINK (http://auth.ohiolink.edu)*, a journal database by the Ohio Library and Information Network. OhioLINK provides a single digital platform of nearly 6,000 full-text journals for more than 600,000 people (*Maximising Library Investments in Digital Collections through Better Data Gathering and Analyses*, Institute of Museum and Library Studies, 2005-2007);

– *ScienceDirect (http://www.info.sciencedirect.com)*, a database of nearly 2,000 journals published by Elsevier and noted for its scientific coverage (*Core Scholarly Research Trends Study: Deep Log Analysis of Elsevier ScienceDirect Users*, Elsevier, 2005);

– *Oxford Journals (http://www.oxfordjournals.org)*, focussing on open access journals published by Oxford University Press (*Characterizing Open Access Journal Users and Establishing their Information Seeking Behavior Using Deep Log Analysis*, OUP, 2005-2006).

The most significant data from these studies has been pooled to provide the deep, comprehensive and robust descriptions of digital library use and users that follows. We have never had such a large data set of usage data. In some cases these data have shocked both publishers and librarians alike, so much so that one librarian, after learning about CIBER findings, was forced to admit that "I don't recognize the users you are describing". This is because they had never witnessed the behavior of so many people before, and never in so much detail.

13.3. Use of digital journals

Everyone readily accepts that the migration of journals to a digital environment has led to a vast increase in the number of journals that an individual scholar can access, usually as a result of Big Deals negotiated by library consortia. The figures are indeed impressive; thus, every user of OhioLINK, the very first Big Deal, has access to the full text of nearly 6,000 journals from every subject field under the sun, some with archives that stretch back 20 years [NIC 06c]. However, there has been some doubt as to whether all this material is actually used [NIC 06d]. Sceptics believe that with the Big Deals (and the like) there is a good deal of wastage and that the availability of such huge digital resources to the user inevitably means that a large number of these titles, and many of their articles, go unused, or are little used. They believe the model is more push than pull, with authors (and publishers) involved in publishing for publishing sake on a grand scale, with readers being overawed or showing little interest in much of the product. Of course, there is nothing new in this and the situation existed in the previous print-only world, but evidence was less easy to obtain then and clearly we now have a difference of scale.

However, all our data show phenomenal levels of use and users, and that the level of use is increasing by leaps and bounds. Thus, take the case of Synergy, a largely scientific, medical and social science database. In 2004, half a million people used the site during a single month and viewed nearly 5 million journal items – and these were deep log figures not the inflated headline hits figures that tend to do the rounds [NIC 05a]. However, these are, of course, aggregated data – what of individual journal titles and articles? To answer that, we shall turn to OhioLINK data [NIC 05b]. From a collection of 5,872 journals available in the month surveyed (October 2004) 5,868 journals were viewed, if content lists, abstracts and articles are included; 5,509 journals were viewed if only abstracts and articles are included and 5,193 journals were viewed if only articles are included. Thus, it would appear that virtually everything on offer was used (99% in fact). This has to be a very powerful statistic – after all only a month's worth of data was involved – a testament to the consumer attractiveness of such enormous digital libraries and Big Deals.

Ohio journals were well used and use was well distributed across journal titles [NIC 06e]. Thus, only about 4% of journals were used once, 9% of journals were used 2 to 4 times, 13% of journals were used 5 to 10 times, 15% 11 to 20 times and 59% 21 or more times.

Nevertheless, there was the inevitable concentration in use. Thus, 5% of journals accounted for well over a third (38%) of usage, 10% accounted for approximately half of usage (53%), and half of all journals accounted for about 93% of use. We can conclude from this that all journals were used, but, in the case of a high proportion of them, not a lot.

We can look at journal utilization another way by examining the number of journals viewed during a search session. This is a very pertinent metric because massive databases, like OhioLINK, are all about massive choice and it is interesting to see what transpires as a result of being offered this. Of the sessions viewing journal items, half only viewed item(s) from one journal, 30% viewed 2 to 3 journals, 14% viewed 4 to 9 journals and 7% of sessions viewed 10 or more journals. With more than one in five sessions viewing more than four journals a good proportion of users do appear to be taking full advantage of the choice offered.

13.3.1. *Downloads (ranked lists)*

Full-text downloads are said to offer the ultimate evidence of user satisfaction and here we find that approximately half of all the Ohio sessions saw a full text article viewed or downloaded and the average number of articles viewed in a session was just over 2. This constitutes a relatively high hit rate given the nature of much searching (a few terms tossed into a search engine).

The number of downloads for top individual journals is very impressive indeed, as Table 13.1 discloses [NIC 05b].

Title	Article views
Journal of the American Chemical Society	48,295
The Lancet	25,046
Journal of Personality and Social Psychology	23,145
Journal of Advanced Nursing	20,958
Applied Physics Letters 1	19,834
American Psychologist	19,370
Angewandte Chemie International Edition	18,765
Journal of Organic Chemistry	18,351
Polymer	17,293
Biochemistry	17,213

Table 13.1. *OhioLINK Top 10 journals by article usage, including vendor traffic (2004)*

13.3.2. *Article use*

An exploratory article level analysis of Ohio journals is particularly illuminating [NIC 06e]. Thus, during October 2004 it was estimated that 319,049 full text articles

were viewed and of these 247,612 were different articles. The study found no evidence that use was dominated by a small number of articles – a common misconception. The top 25 articles viewed accounted for just over half a percent of article views (0.64%). The most popular paper accounted for 0.14% of article views. Most articles were only viewed once. Nearly three-quarters (72%) of articles were viewed once, 25% were viewed between 2 to 4 times, about one and a half percent were viewed between 5 to 10 times and just half a percent of articles were viewed 11 times or more.

13.4. Site penetration and "bouncing"

So far so good regarding the evidence on the attractiveness of digital libraries, but, possibly, on the debit side CIBER data does show that a relatively high proportion of scholars do not make extensive use of the resource, they tend to bounce in and quickly bounce out again. Thus, a characteristic of all our studies (Table 13.2) is that users tend not to penetrate a website site very deeply; typically two thirds were content to examine 1-3 journal items (e.g. an abstract, contents page, full-text document). This could be represented as: a) users demonstrating a lack of interest (or disappointment) in what is on offer, a case, perhaps, of ubiquitous search engine searching throwing up a lot of irrelevant leads; b) a form of direct, fast and assertive searching by which users quickly identify (or already know) what they want; c) the direct result of the massive choice on offer with people quickly moving on to another (more) promising site. It is our view that the apparent shallow searching behavior that is so characteristic of the virtual scholar indicates a checking, comparing and dipping sort of behavior that is a result of easy access, a shortage of time and huge digital choice. It can be equated to the channel hopping behavior exhibited by so many television viewers armed with their remotes. There is a sense that people are increasingly seeking information horizontally and they clearly need a lot of titles to feed off as part of this process; they cross-check information, so a certain amount of so-called redundancy or noise is probably a necessary element of the process.

Type of user/session	Number of items viewed	Emerald Insight (Jan.–Dec. 2002)	Blackwell Synergy (Feb. 2004)
Bouncer/checker	1 to 3	70	67
Moderately engaged	4 to 10	20	26
Engaged	11 to 20	6	5
Seriously engaged	Over 21	4	2
	Total	100	100

Table 13.2. *Digital library – site penetration*

13.4.1. *Infrequent visitors*

Virtual scholars not only tend not to penetrate a digital library very deeply but they also tend not to visit very often, something which has led us to refer to many virtual scholars as being "promiscuous". Indeed, in the case of the EmeraldInsight study, two-thirds of people did not come back within the year [NIC 05a]. This is thought to be another manifestation of massive consumer choice and a dependence on search engines.

Certain types of users were more likely to return than others and Table 13.3 shows that, in a study of Synergy, scientists were more likely to return. The term "academic" in the table refers to social science and humanities users. In the same study it was found that younger scholars were also less likely to return and professors more likely to return.
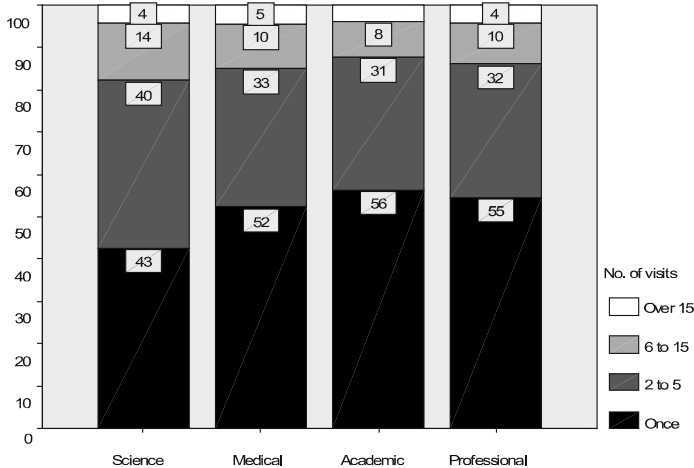

Table 13.3. *Return visits to Synergy by broad subject background of user*

13.5. Reflections on what constitutes a digital library "user"

Thus, on the whole, the CIBER data generally show impressive, albeit uneven, use. However, before we rush to any firm conclusions, we need to reflect on use and users in a digital environment, because all is not quite what it seems. First of all, many so-called "users" were nothing other than robots, spiders and agents, hunting out information for search engines, etc. Thus, in the case of OUP's Nucleic Acids Research, 42% of usage in 2004 was accounted for by these digital entities [NIC 06f]. However, we cannot remove them from the reckoning as clearly they direct users to the site and thus increase human usage as a result of their visits.

13.6. Reflecting on the meaning of "use"

People might view or download a large number of articles but how do we know they actually read or consume them? In fact, a lot of squirreling away does go on in cyberspace, where people are simply gathering articles that they hope to have time to read later or might need some time in the future, but probably never do. To a certain extent we can establish the truth of this. Thus, during one of our studies, Emerald users were offered free access to two journals each week and use rocketed during those two weeks and rose to 10 times the pre-offer figure [NIC 03]. Soundings of the amount of time they spent online, while they were viewing the text, suggested that users were viewing the article much more quickly than was typical in pre-offer weeks (suggesting they were laying them aside for another time). What we were in fact witnessing was a kind of scholarly sales frenzy.

More was learnt about the reading behavior of users of digital journals from the ScienceDirect study [NIC 06g]. It was discovered that the shorter the article the more likely people were to read it online and the longer the article the more likely it was to be read in abstract form only. In other words, if as an author or publisher you actually want people to read something online, then and there, it has to be short.

13.7. Widespread popular interest in digital journals

Another testament to the popularity of the digital journal is the fact that they are very attractive to non-subscribers (i.e. people who do not have full-text rights courtesy of their library, unless they pay with their credit card of course). Indeed, in the study of EmeraldInsight two-thirds of visitors were non-subscribers, seemingly quite happy to rummage around abstracts, tables of contents and what free full-text crumbs were on offer. Thus, many more people are now accessing scholarly journals, and many of these people are novice, occasional users and do not have full text downloading rights. We call these people the "disenfranchised", the type of people who it is argued would benefit from the open access (OA) publishing model. Clearly, the market for scholarly journals has been massively expanded as a result of their digital transition. Obviously, the OA publishing model is attempting to cater for their needs and there must also be (welcome) opportunities for the document delivery community here.

13.7.1. *The rising popularity of the e-journal*

Digital journal libraries are very popular and they are becoming increasingly popular. There has been a huge growth in downloads over the past few years. This in the case of the evaluation of OUP Open, the *Nucleic Acids Research* journal saw

full-text downloads increase by 20% between 2002 and 2003 and by a whopping 213% in 2004. In 2005, article views increased by another 8%; possibly, saturation point being achieved. Thus, over the survey period article downloads increased from 150,000 to 500,000 – some increase.

13.7.2. *Abstracts make a come back*

With the advent of full-text digital libraries it was thought that this would spell death for the ubiquitous abstract, but in fact this has led to a huge increase in the use of abstracts. This result has been a general feature of all our studies but was most marked in the case of the evaluation of OUP Open, where the *Nucleic Acids Research* journal saw an increase in abstract views of the order of 96% between 2002 and 2003 and then jumped 330% in the year that the journal went to open access (2004). The reason for the increasing popularity of abstracts is that they offer a quick and effective means of checking relevance and comparing articles from the massive choice on offer. In some cases they also offer a suitable substitute to the article itself, and this happens particularly in the case of physics, where abstracts are particularly informative. Interestingly, the use of abstracts, except for the age group 26 to 35, tended to increase markedly with the age of the author.

13.8. Search approaches

OhioLINK with its large number of journals and historical archive offered an immense and rich canvass upon which to investigate the methods academics used to find information. The search engine approach proved the most popular by far. Nearly two-thirds of search sessions saw the search engine being used, sometimes in conjunction with other methods. There were big differences in the outcomes resulting from searching and browsing. Thus, by using the alphabetical or journal lists to find information (browsing), users were led to a list of articles and abstracts relating to one issue of one journal. Contrast this to what those people opting for the search engine found. By entering a keyword(s), they were presented with a very different set of choices. OhioLINK delivered to the user a list of article and abstract options irrespective of journal title or issue in response to a user's search query. In consequence, the search option led logically to a greater visibility of both older items and to a wider range of journals and subjects. This is, indeed, what was found. Those people using the search engine were far more likely to conduct a session that included a view to an old article. They were also more likely to view more journals, but not just more journals – they also viewed more articles and abstracts too. There were session time differences too. Those people using the search engine, especially in combination with other access methods, recorded the longest average (median) session times of about 20 minutes and those using the subject lists recorded the

shortest session times. Search-based sessions clearly have different characteristics compared to sessions where users access content by an alphabetical or journal list. Whether someone at a point in time is fact checking – know what they are looking for, or conducting a current awareness search, where they do not know what they are looking for but have a knowledge of the literature – will impact here.

13.9. User diversity

All our work has confirmed that there are very real differences between various types of user, especially in regard to their subject field; academic status and geographical location. This highlights the great danger of trying to generalize data on the back of hundreds of thousands of users (and lumping Nobel Prize winners and undergraduates together as a result) and points to the need to view usage data in the context of user data. We have pointed to some differences already but the sheer diversity in use is best illustrated by referring to the results of the investigation of the use and users of ScienceDirect. Thus, in regard to:

– *Full-text use*. Unsurprisingly perhaps, students made the greatest use of full text (HTML) and less surprisingly perhaps Chinese users recorded the highest relative use of PDFs.

– *Articles in press*. Users from the fields of business and management, engineering and mathematics made higher than expected views of articles in press (AIP). The younger the user, the more likely they were to use articles in press. In regard to occupational status and geographical location, respectively, post/graduates and Japanese users recorded the highest use of AIPs.

– *Successful searching*. Over two-thirds of searches conducted by respondents over 65 recorded zero returns; a case of failure at the terminal? The likelihood of undertaking sessions where only one search was featured increased with age.

– *Obsolescence/decay*. The year in which an article/abstract was published was grouped: current (2004–2005); declining (1999–2003); and old (1993–1998). Older and younger users were more likely to view current material. Users from Spain and China recorded the highest percentage views to current material. In terms of session metrics, business studies users were most interested in current material and hospital staff were most likely to conduct sessions that only viewed current material. Those aged 46 to 55 appeared most interested in old material. Students were more likely to only view current material in a session. There is evidence that those using a search option rather than a menu (browsing) option to locate material were less likely to view current material and were more likely to view historical material. This was because the search engine resulted in the greater visibility of historical material.

– *Viewing most journals.* Those aged 46 to 55 were most likely to view two or more journals in a session – a third of this age group did so. Graduates were most likely to only view one journal, while researchers (staff) were most likely to view two or more journals (30%).

Most digital libraries have a global audience but we should not make the mistake in thinking that every nationality uses the resource in the same way. Thus:

– German authors proved the most active online: they conducted the most searches, viewed the most journals during a search, and made the most requests. Chinese authors were not that far behind in regard to levels of activity;

– Chinese authors were the most likely to view in PDF and Japanese authors were the least likely to;

– French authors were most likely to view older articles and Spanish authors were most likely to view current articles;

– Americans authors appeared to be the least successful searchers and, of course, German authors were the most successful searchers.

13.10. Conclusions

The findings that really stand out from the virtual scholar research were: 1) the massive popularity of scholarly digital journal libraries with the population at large – virtually all of what was offered was consumed in some form or another; 2) the popularity of search engines for locating content and this has led to a very different way of viewing the scholarly information domain; 3) the rejuvenation and increased popularity of abstracts, despite (or because of) almost unlimited full-text access; 4) the "horizontal" form of scholarly information seeking today, a function of unlimited choice, shopping around, short memory spans and search engine searching; 5) the increased visibility of older material in the digital environment which leads to their greater use – something which should lead to a reconsideration of theories of article decay, obsolescence, etc.

We now have a huge evidence base concerning the use and users of digital libraries, especially in regard to digital journal libraries. We are beginning to move away from hits and moving towards the user, and not before time. This had to be the first challenge, now the second challenge is to employ this data to drive information policies and digital library design and development in the virtual scholar field. This is being achieved to a limited extent and CIBER has, for instance, been commissioned to look at the impact of open access publishing on usage. There has been some progress regarding the standardization and utility of usage data thanks to COUNTER (www.projectcounter.org/), although the data produced is rather thin

and anonymous, which limits its value somewhat. However, despite these initiatives and all the good "user" words bandied around at conferences and meetings, digital library developments tend to be system driven and the user is often an afterthought. This was probably only acceptable when it was difficult to get hold of first class data on the user but thanks to deep log analysis, this is no longer the case. It has to be recognized that when a complex and global product, such as a digital library is rolled out, inevitably, things will not go as planned. Therefore, evaluations of the kind outlined here are at the very heart of the design process; they are integral and not an add-on, something nice if you have the time. We should be using deep log analysis to make sure the digital library is always responsive to the users, as new users come into the system and as the information needs of existing users change.

13.11. Bibliography

[NIC 03] NICHOLAS D, HUNTINGTON P. and WATKINSON A., "Digital journals, big deals and online searching behaviour: a pilot study", *Aslib Proceedings*, 55 (1/2), 2003, pp. 84-109.

[NIC 05a] NICHOLAS D., HUNTINGTON P. and WATKINSON A., "Scholarly journal usage: the results of deep log analysis", *Journal of Documentation* 61(2), 2005, pp. 248-280.

[NIC 05b] NICHOLAS D., HUNTINGTON P., RUSSELL B., WATKINSON A., HAMID R., JAMALI HR. and TENOPIR C., "The big deal: ten years on", *Learned Information*, 18(4), October 2005, pp. 251-257.

[NIC 06a] NICHOLAS D., HUNTINGTON P., JAMALI HR. and WATKINSON A., "The information seeking behaviour of the users of digital scholarly journals", *Information Processing & Management*, 42(5) pp. 1345-1356, 2006.

[NIC 06b] NICHOLAS D., HUNTINGTON P., and JAMALI HR., "Authors as users: a deep log analysis linking demographic and attitudinal data obtained from Elsevier authors with their usage of ScienceDirect", London: UCL, 2006 [unpublished].

[NIC 06c] NICHOLAS D., HUNTINGTON P., JAMALI HR. and TENOPIR C., "OhioLINK – ten years on: what deep log analysis tells us about the impact of Big Deals", *Journal of Documentation,* 62, 2006.

[NIC 06d] NICHOLAS D. and HUNTINGTON P., "Digital journals: are they *really* used?", *Interlending and Document Supply*, vol. 34(2), 2006.

[NIC 06e] NICHOLAS D. and HUNTINGTON P., "Ohio Link Briefing paper 6", London: UCL, 2006 [unpublished report].

[NIC 06f] NICHOLAS D. and HUNTINGTON P., "Characterising Open Access journal users and establishing their information seeking behaviour using deep log analysis techniques: case study Oxford Open", London: UCL, 2006 [unpublished report].

[NIC 06g] NICHOLAS D., HUNTINGTON P. and JAMALI H., "Obtaining rich use and user data on virtual scholars: a deep log analysis of ScienceDirect authors", London: UCL, 2006 [unpublished paper].

Chapter 14

Digital Collections in Libraries: Development and Continuity

14.1. Introduction

Libraries are keepers of memory – "the memory of mankind", as Johann Wolfgang von Goethe affirmed. They rely on the memory of their past to better live the present and prepare for the future. Through not decades but centuries and millennia of existence, libraries have transformed themselves and, in particular, have seen changes in the documents they gather and protect, to better promote their use. We have seen clay tablets, papyrus scrolls, manuscripts, chained books, printed texts multiplied in the form of books, journals, magazines and newspapers, and, more recently, microforms and audiovisual documents, which have also had their day. At the turn of the 21st century, digital documents have incontestably made their entrance on library "shelves".

Libraries know how to evolve, using the inventions and the technologies, always towards the goal of accompanying, even anticipating, the needs of the clients they serve. Already, even though we have not come all the way through every stage of the process and obtained mature results, digital resources are present in libraries of all types: large and small, specialized and general.

Digital documents support texts, images and sounds as did their predecessors: print, microform and audiovisual documents. It is again the medium that is transforming itself, drawing on the information and telecommunications

technologies – often still called "new", even though they have been with us for a few decades now.

With digital resources, it is not just the medium that is changing. New ways of organizing content and using it, enabled by the same technologies, constantly afford new possibilities. In fact, the changes in and new potentials for documentary products are often triggered by the latent or expressed needs of libraries and library users – always hungry for more. In this way, electronic resources generate multiple adaptations and affect all stages of the document chain in libraries: from identification, through acquisition, intellectual and physical processing, storage, circulation and preservation, to weeding.

Digital documents raise a new challenge for libraries and their staff, as well as documentation and information specialists: the harmonious and successful integration of digital resources into the holdings and collections of libraries and other documentary institutions. The adaptation required affects activities not only in documentary processing and collections management, but also in the delivery of digital resources to the targeted clientele.

14.2. Adaptations and alterations in the document chain

The adaptation of the library workflow in order to integrate the digital resources speaks for the ability libraries have to evolve. The digital documents bring changes that affect more than the physical format. They impact the organization and use of the library content, these impacts made possible by the technologies themselves. The changes – without a doubt profound ones – engendered for publishers, producers, exhibitors and booksellers, and, of course, the authors and creators at the start of the chain will not be dealt with here. Those problems probably constitute one of the factors hampering original digital creation, because the issues are complex and fraught with consequences in terms of role-sharing and of intellectual and commercial copyright. This article will focus instead on the changes experienced by libraries from the moment a document exists – when it has been produced and can be "acquired" by a library for its users.

14.2.1. *Identification and selection*

The identification of documents of likely interest to a single library continues to draw on the usual methods of discovery and watch: reporting by "publishers", advertising, reviews, specialized tools for the selection of documentary resources, magazines and scientific articles, colleagues, etc. To these standard modes of discovery have been added not only new sources of information, but, above all, original work techniques and reporting tools. An example is the computerized

reporting of new content in, and updates of, Internet electronic resources made possible by specialized tools or, more recently, by RSS (really simple syndication).

Historically, libraries have faced the problem of deciding whether or not to acquire compilations of works, either in addition to separate editions of the same works already in their possession or in order to fill gaps in their holdings. However, today the grouping of titles is unusual in scope and requires analyses of clearly greater complexity. For example, many databases of journals and magazines propose hundreds, even thousands, of titles that are also found, in significant proportions, in other databases. If budgets are to be used wisely, the people in charge of selection and acquisition must carefully compare the products available in terms of their general interest, originality and overlap. Given the amount of money involved in purchasing a database, all due attention to this process is worthwhile and justified. To make the selection process even more complex, the content of a database varies during the subscription period. Thus, the process must be repeated periodically in order to decide whether to renew a resource or replace it with a different one. Then there is the question of the stability and persistence of a title in a library's collection: just as libraries have not been in the habit of subscribing to a magazine one year and abandoning it the next, they must not unduly disrupt database licenses.

The grouping of titles is no longer limited to databases of journals, magazines and newspapers. The same thing is happening with monographs. Many suppliers now offer vast selections of electronic books, grouped by theme, for instance. While avoiding the trap of examining each title included in a given set, which would be inefficient, it is nevertheless appropriate to examine the titles contained in a supplier-determined grouping enough to assess how worthwhile they are overall given the established price, which will remain to be negotiated. This brings us to the next question: negotiating prices and rates.

14.2.2. *Purchases, subscriptions and licenses*

The electronic resources "market" does not mirror the book market, at least not yet or to no great extent. However, with books, a large number of titles are available "in all good bookstores", and there is still little duplication of titles between "retailers" of electronic documents. Here, exclusive or near-exclusive distribution is much more the rule than the exception.

In the absence of competing suppliers, the prices asked are less easy to validate and evaluate. They are rarely made public or displayed clearly for purchasers. Negotiation is generally inevitable and can rarely be engaged in once and for all with a given supplier. In general, the price of each documentary resource or product must be negotiated separately.

Another major difference with electronic resources is related to the terms of "acquisition" of content. Whereas, historically, a library would purchase a book, magazine or video and possess it for an unlimited time, now it must decide what kind of purchase to make. Does it want to acquire the resource or subscribe to it? Does it want to subscribe for a year or for a longer period? Does it want to pay for a period of use corresponding to a license term or pay for a long-term or unlimited right of use?

To summarize, the decisions relating to the acquisition of electronic resources are often more complex and significant in budget terms than book purchases are. Regarding the latter, a title chosen by a Quebec public library, for instance, costs, on average, about $30, which bears no comparison with the costs of electronic resources totaling hundreds or thousands of dollars. Also in Quebec, the *Act respecting the development of Quebec firms in the book industry (LRQ, c. D-8.1)* governs purchases of books by subsidized libraries: purchases must be made in an accredited bookstore, which simplifies the process of selecting suppliers, and prices are set by a regulation under the same Act (negotiation is not allowed). The acquisition of electronic resources greatly modifies the steps that Quebec's public libraries and school libraries have to take.

14.2.2.1. *Consortium, consortia and company*

In this context of complexity and large-scale costs, variable rate schedules and inevitable negotiations, libraries willingly join together and form purchasing consortia to share in the acquisition of the electronic resources they wish to offer their users. These group efforts enable libraries to think collectively about the interest and value of existing electronic resources, as well as purchasing conditions and terms of access by library users and staff.

A number of consortia have been formed, and libraries often join more than one. Many consortia bring together libraries serving the same kinds of clienteles: public libraries, university libraries, etc. Others acquire resources of common interest for all the libraries in a given territory, in Canadian provinces for example, as with group purchases in Ontario (Knowledge Ontario), Alberta (Alberta Library) and British Columbia (British Columbia Electronic Library Network). Sometimes, too, agreements between consortia make it possible to use the power of a group to facilitate the purchase of a resource that is significant, but of only marginal interest to a given group. For example, it is more difficult for a group of libraries serving a primarily English-speaking population to negotiate for a resource in French, and vice versa. Thus, agreements are entered into whereby negotiations for a resource with French content are entrusted to a consortium serving a primarily French-speaking population, and access for minority Francophone populations in interested Canadian provinces is provided for. Groups of English-language libraries can render

a similar service to French-language libraries. Everyone benefits: reduced overall effort, advantageous prices, improved access. It is important to point out that suppliers also benefit from such arrangements.

14.2.3. *Intellectual and physical processing of documents*

Before they can be made accessible online, the digital documents must follow the necessary steps of cataloging, indexing and classification. Document processing contributes to the fundamental identity of libraries, namely, "organized sets" of "selected resources" to properly serve users. The digital resources at the same time gain from this strength of the libraries and initiate changes in the library workflow.

14.2.3.1. *Cataloging and enriched cataloging records*

Intellectual processing is perhaps where the adjustments required by electronic resources have been the easiest to make. Cataloging standards are long-established and mature. Apart from a few adjustments inherent in the specific characteristics of electronic resources (in the physical description area for an item, for instance), the *Anglo-American Cataloging Rules*, widely used in Québecois, Canadian and American libraries, allow electronic resources to be processed with no great difficulty. These revised Rules cover the cataloging of the digital resources: a specific chapter is dedicated to them, included in a code serving the needs of all types of documents. Soon, the digital resources will also find their place in the new edition, Resource Description and Access (RDA), proud descendant of Panizzi (1839), Jewett (1853) and Cutter (1876).

In addition, the new technologies allow cataloging records to be enriched with new elements, this time for books and other traditional documents as well as for electronic resources: thumbnails of cover pages, tables of contents that are very useful for identifying and assessing the relevance of a document, and links to related resources on the Internet, such as reviews of the title described, etc.

14.2.3.2. *Classification*

In the area of classification, it might seem superfluous to assign a class number to electronic resources, given that they do not need to have call numbers and take their places beside other titles on library shelves. As testified by concrete applications and research[1], assigning such numbers to them means that an expert classification system can be used to find resources dealing with the same subject. This provides a worthwhile complement to the indexing system (keywords, subject headings, etc.) widely used by reference librarians to guide users and by seasoned

---

1 See the studies of Hiskey [HIC 1999] and Normore [NOR 2000].

library users themselves. The latter understood long ago that, with the call number of just one document that meets their needs, they can browse the shelves and find many others of interest that are surely available. Call numbers enable browsing by subject among electronic resources, as many examples bear out. In French, this functionality can be seen at work in *BREF*, the online reference library of Bibliothèque et Archives nationales du Québec (http://services.banq.qc.ca/sdx/bref).

Once again, the genius of Melvil Dewey, the creator of the *Dewey Decimal Classification (DDC)* system, transcends the ages and can be used to advantage with electronic resources. In his biography of Dewey, Dawe reports, "As Dewey explained, it made more sense for a library 'to assign numbers to its books, which are permanent, and not to its shelves, which are liable to frequent changes'".[2] Classification systems, such as DDC or the Library of Congress Classification, remain useful for classifying digital resources. Come to think of it, is it not considerations similar to those evoked by Melvil Dewey that lead computer specialists to design databases with permanent addresses for data? Similarly, just a few years after the Internet came into existence, PURLs ("Persistent Uniform Resource Locators")[3] were born to deal with the difficulties related to the constant mobility of Internet resources. As long as a performing searching system, allowing different degrees of specialization, has not entered the libraries and their integrated systems, the classifying of documents remain useful and justified.

14.2.3.3. *Indexing and full-text searching*

It might be thought that full-text searching has replaced document indexing and rendered it obsolete: "the greater contains the less", as the saying goes. This could be countered with "enough is enough", speaking of the over-abundant results obtained with a Google-type search, when a search engine is used to query huge sets of information or when a search term is not specific enough for the documentary holding queried.

The main advantage of using indexing is that documents more closely related to a search topic can be found more quickly, especially when the search terms cover a large number of documents in the target queried. If there are no results, it is never too late to extend the query to full texts.

Indexing limited to the three or four main subjects dealt with in a specific resource allows documents meeting a specific need to be retrieved more easily and surely. In this area, search engines working with full texts have not yet attained the degree of refinement required to adequately weigh the interest of a monograph devoting an entire chapter to a given subject against that of a magazine article

---

2 Dawe, *Melvil Dewey*, p. 174.
3 http://purl.oclc.org.

dealing with the same subject in a single, superficial paragraph. Of course, this is not to deny that a recently published in-depth journal article on a specific question can adequately meet a user's needs. However, the problem of weighting search results on the basis of terms – whether controlled or natural-language terms, whether chosen by an indexer or appearing in the texts themselves – is still far from resolved. This chapter will not devote the attention required to such a complex subject. Improvements in search engines in the coming years will surely provide ways of addressing, at least in part, these interesting problems of information searching and retrieval.

14.3. Searching and catalogs …

Finding information or a document, like the discovery of reading for pleasure, can come about in a great variety of ways. A work may be reviewed or cited in a journal, magazine or newspaper, or a friend, peer or bookseller may make a recommendation – to give just a few examples not involving the riches of libraries. To those who use them on site or remotely, the libraries offer catalogs, imperfect instruments certainly, but proven and devised to answer users' questions or librarians helping users.

It is not surprising that libraries have relied on the strength of their document processing and catalogs to report, describe, organize and provide access to electronic resources, alongside the other documentary resources. In a short time, cataloging standards (MARC, Dublin Core, OAI-PMH protocol (Open Archives Initiative Protocol for Metadata Harvesting), etc.) and integrated document management software have made it possible to link the bibliographic records in catalogs to the computer files containing the digital resources themselves, providing for immediate on-screen display not only of the citation of a resource, but also of its full content.

Before turning to the impacts of electronic resources on library catalogs, it is interesting to note that the information technologies have given a second or third or umpteenth life, even more useful than the previous ones, to library catalogs. Querying them is clearly easier than searching the card catalogs that preceded them – even user-friendly, especially if a user is "computer-literate". As a result, today, the computer stations made available to the public in libraries are heavily used to query the local catalog. Additional convincing evidence of catalog use is provided by the visitor statistics of library Internet sites and portals, which often show the catalog at the top of the sections most visited. The arrival of the Internet and search engines and the use of information technology and computers in various spheres of

activity (bank transactions, work, video games, etc.) surely contribute to making the public more comfortable with computer equipment and, consequently, with computerized library catalogs.

Nevertheless, despite their long history and the immense efforts devoted to their development, computerized library catalogs still pose undeniable problems, some related to their lack of user-friendliness, others to the insufficient skills of users.

In addition, having helped resolve certain historical difficulties of library catalogs (establishment, maintenance, accessibility and data sharing), the information technologies, because of their immense, ever-increasing capacities, have themselves generated new problems, notably those related to the birth of mega-catalogs.

14.4. … searching and mega-catalogs

Mega-catalogs now contain resources of every sort: monographs; periodical articles; archival and museum resources; electronic resources provided locally, on the Internet or through external providers, etc. This produces huge catalogs for institutions with very rich holdings. In this area, as in so many others, globalization is at work: the WorldCat catalog of the Online Computer Library Center (OCLC) contains more than 82 million bibliographical records, one billion items from over 10,000 libraries.

This immense scale unfortunately has its downside. A search often leads to over-abundant results that are difficult to rank by relevance with respect to need. The solutions proposed to this problem by commercial search engines (Google, Yahoo, etc.) are not always in line with the objectives pursued by libraries, to wit: the best possible response to a need for information, clearly independent of commercial objectives and, within certain limits, dissociated from factors like "the most popular titles".

For a number of years, studies have been conducted on this question[6], and developments have been added to the search modules of software and document management systems to improve the results of searches in catalogs. Work has been done on data aggregation in order to group together, for instance, the various manifestations of a work, in its diverse expressions: a printed book, a musical work, a cinematographic work or a digital reproduction of any of the three. For an experienced cataloger, this evokes the problematic "uniform titles" of cataloging

---

6 Sleeman [SLE 2005].

rules that were aimed precisely at grouping documentary resources together under a unifying access point. This brings us back to "development and continuity".

No sooner have a few solutions for mega-catalogs been thought up and implemented than complexity rears its head again, notably with federated search engines. These types of engines simultaneously query the catalogs of several libraries, to which may be added neighboring sources of information including museums, archives centers and more. New tools can be used to query a large number of sources, including, for instance, electronic periodicals with full texts of articles such as the product *SerialsSolutions* or integrated functions of library systems.

14.5. Organization of collections

Searching also works with the organization of collections, as it does with Internet site plans. Even at the modest scale of a library portal, the ideal model for the organization of electronic resources, on a par with the organization of books on library shelves, does not yet exist, at least to my knowledge. Traditionally, in libraries, the spatial organization of books and other documents complements the results obtained with indexing and classification systems. Larger categories, by field (e.g. science and technology, humanities and social sciences, arts and literature, etc.) as opposed to broader categories (fiction and non-fiction), or by medium (e.g. microforms) or type of document (e.g. comic books, audio books, journals and magazines), are often used in libraries to guide users.

14.6. Physical processing, accessibility and placement online

The physical processing of electronic resources, which is far from virtual processing, requires due attention, albeit of a different type. Here, the binding of books or issues of a journal or the boxing of an audio-visual document give way to the operations of saving and storing electronic resources, which also generate metadata to be established, preserved and updated. While the life expectancy of a bound book can readily be estimated at a few hundred years, electronic formats do not promise similar stability: we must see to preserving electronic resources in a format that can continue to be "read" and disseminated. This implies a new focus and leads library staff to acquire new knowledge and work in a multidisciplinary way that combines library science, information science, and information and communications technology.

Before deciding to purchase an electronic resource, a library will have to check what software and hardware is required to exploit and disseminate it. This step must occur before a purchase or subscription is confirmed. Again, this is very different

from the situation in which a book made its way to the library shelves, joining all those already there without further ado, ready to be devoured by an avid reader.

Again, we have development and continuity: processing and preserving the better to provide access, in the short, medium and long term. Clearly improved services, holdings extending beyond the walls of institutions and geographical boundaries, and efficiency that would have been unimaginable just a short while ago – all are significant improvements we no longer want to do without, but they require adapted budgets, which cannot be presumed to be lower, as I am sure you have guessed.

14.7. Preservation …

The preservation of electronic resources seems to lend itself to dialog and sharing more easily than the printed or audiovisual documents. With books and other print materials, numerous attempts, sometimes requiring decades of discussions, have generated a very limited number of common storerooms or deposit libraries. On the other hand, with electronic resources, a number of consultation and cooperation projects are already coming up with common shared solutions to the problem of long-term preservation[7].

Naturally, libraries have not resolved all the problems in this area and they will have to negotiate a great deal yet, first among themselves and then with commercial information providers, in order to keep perennial access to electronic resources. From the viewpoint of libraries, this is amply justified by the often high costs they pay, particularly for licenses. Democratic access to knowledge for future generations depends on it.

14.8. … and dissemination

The difficulties linked to the persistent access to digital resources have even led to a reversal of situation: libraries and institutions are now becoming producers of information. For example, to guarantee true access to their clientele, unlimited in time, and at a reasonable cost, universities combine their efforts so that, from now on, they can themselves publish the research results of professors and researchers. Institutions have become electronic resource publishers. *Érudit*,[8] for instance, in the Quebec environment, provides a place for the dissemination of university journals,

---

7 Deegan [DEE 2006].
8 http://www.erudit.org/info.html.

electronic books and dissertations. The main partners in the project are the University of Montreal, Laval University and the University of Quebec at Montreal.

At the same time, a number of large digitization programs have built imposing digital collections: *Gallica*, from the Bibliothèque nationale de France; *American Memory*, from the Library of Congress; and the more recent *Europeana*, to name but a few. In Quebec, Bibliothèque et Archives nationales du Québec offers abundant, diversified digital resources on its Internet portal: books, maps, sound and video recordings, postcards, adapted documents for visually-impaired persons, private and public archives, genealogical sources, etc. Some of these resources are original and digitally conceived, such as the videos presenting interviews with poets.

Dissemination implies users' training for the discovery of the libraries' richness. Even if a larger number of people know how to use a computer or carry out banking operations, they do not necessarily easily navigate library portals and catalogs. That training task has always been present and the libraries must continue to fulfill it in order to meet their mission. Under those conditions only, the citizens will be able to freely use the libraries, as perfect democratic institutions giving access to knowledge and culture.

14.9. Conclusion

The library workflow is complete: from production to dissemination, by way of research and preservation. Also, at each stage along the way, digital resources impose changes, adaptations and new tools. I fully realize that I have barely touched on several questions that have been investigated in depth by others or that remain to be delved into. A great deal of thought and effort is still required in order to meet people's needs for information, education and culture in the most user-friendly way possible for everyone and thereby contribute to democracy.

With digital resources, we are undergoing many profound changes, as has happened before in history, but beyond the differences in medium, potential, tools to exploit the new resources, and ways of disseminating and preserving them, the mission and objectives of libraries remain fundamentally the same: to put knowledge- and culture-seeking individuals in touch with knowledge sources, and to do so democratically and without discrimination. To borrow the words of the 19th-century German-American political philosopher and jurist Francis Lieber, libraries will thus continue to be "the bridges over which civilization travels from generation to generation and from country to country".[9]

---

9 Quoted in Michael H. Harris [HAR 1995] p. 201.

Libraries have crossed the centuries, adapting themselves and adopting changes, evolving for the greater benefit of their users. They will make their way through this digital era successfully and continue their work. Already, the so frequently asked question, "Do libraries have a future with the coming of the Internet?" has been transformed into "How can it be that libraries are so useful and used so much in the Internet era?". The answer is surely more than just tradition: *development and continuity!*

14.10. Bibliography

[CON 2005] CONNAWAY L., Physical Space for Virtual Services and Collections, Portal: Libraries and the Academy, vol. 5, number 1, January 2005, pp. 127-131.

[DAW 1932] DAWE G., *Melvil Dewey: Seer, Inspirer, Doer, 1851-1931.* Melvil Dewey Biography, Lake Placid Club, NY, 1932.

[DEE 2006] DEEGAN M., TANNER S., *Digital Preservation*, Facet, London, 2006.

[HAR 1995] HARRIS MICHAEL H., *History of Libraries in the Western World,* 4th ed., The Scarecrow Press, Lanham, Md & London, 1995.

[HIC 1999] HICKEY T.B., VIZINE-GOETZ D., "The Role of Classification in CORC", Annual Review of OCLC Research, 1999.

[LER 1998] LERNER F., *The Story of Libraries: From the Invention of Writing to the Computer Age*, Continuum, New York, 1998.

[NOR 2000] NORMORE L., BENDIG M., "Using a Classification-based Information Space", *Workshop on New Paradigns in Information Visualization and Manipulation, in conjunction with the ACM Conference on Information and Knowledge Management,* Washington, DC, November 10-11 2000.

[SLE 2005] SLEEMAN B., BLUH P.M., *From Catalog to Gateway: Charting a Course for Future Access: Briefings from the ALCTS Catalog Form and Function Committee*, Chicago, Association for Library Collections & Technical Services, American Library Association, 2005.

[TOL 2001] TOLZMANN DON H., HESSEL A., PEISS R., *The Memory of Mankink: The Story of Libraries Since the Dawn of History*, Oak Knoll Press, New Castle, Delaware, 2001.

Chapter 15

Ergonomic Standards and the Uses of Digital Libraries

15.1. Introduction

The users of digital libraries want to be able to have easier access to online resources and to be able to access these resources as quickly as possible, however, the access to documentation has become more and more complicated. Although it has become much easier to enter a digital library and become familiar with the classification of the books on the shelves, it has become very difficult to work with the various platforms made available by publishers, library catalogs, specialized databases or portals which cater for the needs of different communities[1]. With this in mind, the user has adapted to only one search engine, Google. Google, with its almost instantaneous responses, responds to the different needs of the users. The interface is extremely basic: the user only needs to type the keywords in the search, or even the question they need answered. All that remains for Google is to make sure of the relevance of the results because the user is faced with thousands of results amongst the billions of indexed pages.

Documentalists thus play a fundamental role in the classification and indexing of documents. Libraries have constructed more precise interfaces on their websites which in turn are more relevant for the users. Which ergonomic standards do the digital libraries rely on? What are digital libraries currently being used for? This chapter is divided into two parts: the first part tries to make a synthesis of the ergonomic standards and style

Chaper written by Nicole LOMPRÉ.

[1] As an example, have a look at the list of web tools on the site of the INIST-CNRS http://bibliostic.inist.fr/ article69.html.

guides which have evolved over time, the second part analyzes the uses of library interfaces and the usability tests which were carried out in libraries in relation to the ergonomic criteria.

15.2. The evolution of ergonomic standards for user interfaces

A lot of ideas on user interfaces were developed or refined at the Xerox Palo Alto Research Center in the 1970s from the following eight principles [CAN 82, BAE 87]:

1) a conceptual model which is familiar to the user, such as the office metaphor,
2) see and point rather than remember and type,
3) see and obtain (WYSIWYG: *What You See Is What You Get*),
4) universal commands,
5) coherence,
6) simplicity,
7) flexible interaction[2],
8) interfaces which can be personalized[3].

In the 1980s the research companies at the French National Institute for Research in Computer Science and Control (Institut National de Recherche en Informatique et Automatique or INRIA), and in particular the research group working on the ergonomic psychology of the design of human-machine interfaces, along with the work of Dominique Scapin [SCA 86], came up with eight ergonomic criteria (each of which is broken down into sub-criteria that will not be dealt with here):

1) compatibility,
2) homogenity,
3) guidance,
4) adaptability,
5) work load,
6) explicit control,
7) error management,
8) meaning of codes.

The work carried out by Joëlle Coutaz (who visited the ergonomic psychology teams at the IINRIA), as well as what she suggests in her book published in 1990 [COU 90] should also be mentioned; she has a vision which is more suited to computer scientists using five criteria:

1) structure of activities,

---

2 *Modeless interaction*.
3 *User tailorability*.

2)   coherence,
3)   flexibility,
4)   feedback,
5)   conciseness.

The work of Jakob Nielsen [NIE 93] suggests five attributes which are linked to usability:

1)   the ease with which users can learn how the system works,
2)   performance of the system in order to have a high level of productivity,
3)   the system can easily store previous searches carried out by users,
4)   the low number of errors,
5)   user satisfaction.

All of this research had a large impact on the French standard number AFNOR Z67-133-1 from December 1991 entitled "The definition of ergonomic criteria for the design and evaluation of user interfaces". This standard number keeps seven of the eight original ergonomic criteria by changing certain terms and by removing the work load criteria which is then divided amongst the other seven:

1)   *compatibility*: getting the users to become more involved in a particular activity, for example the organization of functions in keeping with the task in hand;

2)   *homogenity*: the stability of the design choices, for example the presentation of information. Homogenity also involves using identical procedures which lead to the same result;

3)   *guidance*: the methods made available to the users, such as the structure of the display or messages;

4)   *flexibility*: the interface's capability to adapt to the needs of different users, such as beginner/expert modes or the possibility of defining the parameters or of chaining the frequently used commands;

5)   *explicit control*: methods which enable the user to deal with the effects of the commands, for example signaling the active options in the menus or being able to interrupt a print job;

6)   *error management*: guiding the user in the identification and correction of errors such as the clarity of messages or the command Cancel Typeface;

7)   *conciseness*: the reduction of perception and storing activities, for example with default icons or default options.

Globally, the first standard numbers for the ergonomics of the human-machine interfaces were published from 1992 onwards in a series of 17 standard numbers "9241" which today are used by the majority of ergonomists [AFN 03]. This series

of standard numbers is continually evolving[4]. Several standard numbers are particularly important for ergonomists.

The ISO standard number 9241-11 dating from 1998 defines the usability guidelines as "the degree to which a product can be used by identified users to fulfill the defined objectives satisfactorily while incorporating a high level of performance and within a specified context". In order to measure usability, the following three points must be considered:

1)   describing the desired objectives;
2)   describing the elements of the context for use (users, tasks, equipment and environment);
3)   defining the target values in terms of efficiency, performance and satisfaction [HOR 06] for the desired contexts. This standard number is undoubtedly inseparable from the 12 methods of usability for the user-centered design which are suggested in ISO/TR 16982, published in 2002 and which will become 9241-230.

The ISO standard number 13407, published in 1999 and which is currently being revised under the editions 9241-200 (introduction) and 9241-210, defines the forward planning and management of the user-centered design for the interactive systems. There are four requirements:

1)   the active participation of the users and the clear understanding of the requirements linked to the user and to the task;
2)   the appropriate distribution of functions between the users and the system;
3)   iteration of the design solutions;
4)   multidisciplinary design.

The ISO standard number 9241-10, published in 1996 and which became ISO 9241-110 in 2006, today keeps seven of the principles for dialogs between man and machine:

1)   *adapting to the task:* the dialog makes it possible to carry out a task efficiently and with a high level of performance;
2)   *auto-descriptive character*: the dialog can immediately be understood thanks to feedback from the system or can be explained at the request of the user;
3)   *user control*: the user can initiate and control the direction as well as the pace of the interaction;
4)   *conforming to user expectations*: the dialog is coherent and corresponds to the user's needs and expectations;
5)   *error tolerance:* the predicted result can be obtained either by following the correct procedures or not;

---

4 The entire set of the ISO 9241 series is available at:
http://isotc.iso.org/livelink/livelink?func=ll&objId= 651393&objAction=browse&sort=name.

6) *aptitude for individualization*: the interface can be modified in order to adapt to the task's needs or to the preferences or competences of the users;

7) *facilitating the understanding of how the system works*: the dialog supports and guides the user so they can garner a better understanding of how the system works.

The ISO standard number 9241-12, published in 1998 and which will become 9241-111, also states seven principles for the presentation of information which have been highlighted in the previous version 9241-110:

1) *clarity*: the content is displayed quickly and with precision;
2) *discriminability*: the information can be distinguished with precision;
3) *conciseness*: only the information which is relevant to the task is displayed;
4) *coherence*: the same information is presented in an identical manner on all applications;
5) *detectability*: the information is adequately coded in the right place;
6) *readability*: the information is easy to read;
7) *comprehensibility*: the meaning of terms is easily understood.

A series of three standards from the ISO standard number 14915 which are linked to multimedia (and which are also important in the design of audio-visual objects) states four principles:

1) communication objectives,
2) perception and comprehension,
3) adaptation to exploration,
4) adaptation to commitment.

Beyond these international standard numbers, which are quite general and guide both the designers and evaluators, it must be pointed out that more precise guidelines have been developed. These guidelines detail all the questions and issues relating to the interfaces. Amongst all this work the most interesting, for the moment, is undoubtedly the guide which was created in 2003 for the National Cancer Institute [LEA 06], where the 187 guidelines rely on the results of scientific works. Updated in 2006, the current 209 guidelines are spread over 18 chapters:

1) the design and evaluation process,
2) optimizing the user experience,
3) accessibility,
4) hardware and software,
5) homepage,
6) page layout,
7) navigation,
8) scrolling and page-numbering,
9) headers, titles and sections,

10) links,
11) appearance of the texts,
12) lists,
13) interface controls (objects),
14) graphics, images and multimedia,
15) writing for the Web,
16) organization of the content,
17) research,
18) usability tests.

A synthetic grid groups together the questions at the end of the document.

We can also include the seven "Design for All" principles suggested by the College of Design[5] by North Carolina State University (NCSU) in 1997:

1) *fair use*: the design is used by and marketed to individuals with different capabilities;

2) *flexibility in use*: the design adapts to a range of individual performances and capabilities;

3) *simple and intuitive to use*: the design is easy to understand depending on the experience of the users, their knowledge, their language or their level of concentration;

4) *distinguishable information*: the design transmits relevant information to the users independent of the context or of the sensory capabilities of the users;

5) *error tolerance*: the design minimizes risks, accidents or involuntary actions;

6) *minimum physical effort*: the design can be used efficiently, comfortably and with the minimum of fatigue;

7) *appropriate size and space for approach and use*: appropriate size and space makes it possible to deal with, reach and manipulate the information independent of the size of the user, their posture or their mobility.

This list of principles is becoming significant once again because it includes an e-inclusive and e-accessible vision which is fashionable within current European projects.

15.2.1. *Guidelines for leading computer manufacturers*

Computer manufacturers such as Microsoft, Apple and IBM have also developed their own guidelines which will be explained in more detail below. Other computer manufacturers such as Sun and Linux etc. have developed their own guidelines too.

---

5 *Design for All*: http://www.design.ncsu.edu :8120/cud/pubs_p/pud.htm.

Microsoft[6] includes six principles in its guidelines for user-centered design:

1)   *user control*: the user should always feel in control of the software rather than having the impression of being controlled by the software;

2)   *direct manipulation*: the design enables users to directly manipulate representations of the information provided by the software;

3)   *error tolerance*: the users are able to explore the interface and learn by trial and error;

4)   *feedback*: it is necessary to provide feedback relating to the users' actions;

5)   *aesthetics*: the visual design is an important part of the interface. The visual attributes give a certain impression and enable the user to interact with particular objects;

6)   *simplicity*: an interface must be basic, easy to learn how to use and be easy to use. It must also let the user access all of the application's functions.

When Steven Jobs visited Xerox Park in 1979 he was impressed when he saw the screens and applications in motion. The result was that in 1983 Apple launched Lisa, a product which was similar to Star in terms of the user interface, but with some significant differences [CAN 82].

Apple literally took off in 1986 with its Macintosh and a design guide for interfaces which was revised for its 2006[7] edition with 14 principles:

1)   *using metaphors* which reflect the mentality of the user and which must be a) familiar, b) simple, c) available, d) discoverable;

2)   *explicit and implicit actions*;

3)   *direct manipulation*: leave the objects visible to the eye so they can be worked on;

4)   *user control*: enable the users, and not the computer, to launch and control the actions;

5)   *feedback and communication:* keep the users informed by providing appropriate feedback;

6)   *coherence*: enable the users to transfer their knowledge and habits from one application to the other;

7)   *WYSIWYG (What You See Is What You Get)*: avoid differences in what the users see on the screen and what they receive as an end result;

8)   *error tolerance*: actions are easily reversible which makes it possible to explore the applications;

9)   *perceived stability*: the interface defines standard graphical elements, such as menu bars, window commands etc. in order to create an environment in which the users understand everything, as well as an environment which is both familiar and predictable;

---

6 Microsoft: http://msdn.microsoft.com/library/default.asp?url=/library/en-us/dnwue/html/ch02b.asp.
7 Apple: http://developer.apple.com/documentation/UserExperience/index.html.

10) *aesthetic integrity*: the information is well organized and coherent with the principles of good visual design;

11) *flexibility of the interaction*: let the users do what they want at all times;

12) *managing complexity*: develop an easy-to-use software by keeping the design as simple as possible and close to user needs;

13) to think of the users by providing large-scale compatibility linked to cultural differences;

14) universal access which should be planned at the beginning of the design phase.

IBM[8] states 11 design principles on its website:

1) *simplicity*: do not compromise the use of the different functions;

2) *support*: place the user in control and provide contextual assistance;

3) *familiarity*: create interfaces in relation to the user's expertise and know-how;

4) *evidence*: ensure that the objects and their controls are visible and intuitive;

5) *encouragement*: ensure that the actions can be predicted and can therefore be reversed;

6) *satisfaction*: create a sense of progress and accomplishment;

7) *availability*: make all objects available at all times;

8) *reliability*: avoid creating technical problems for the users;

9) *adaptability*: develop alternative interactive designs;

10) *personalization*: enable the users to personalize their interface;

11) *aesthetics-affinity*: bring objects to life through the use of good visual designs.

15.2.2. *Recommendations by WCAG for accessibility and standard number ISO/DIS 9241-171*

Very recently, the French law from 11th February 2005 (number 2005-102) for the equal rights and opportunities, participation and citizenship of disabled people (which was published in the official journal – edition number 36) stipulates in article 47 that "the services of online public communication provided by state services, local governments and public institutions must be made accessible to disabled people".

This new law will dramatically change the design of the web pages of state services and will also undoubtedly change library interfaces in order to provide access to online documentation for everyone.

---

8 IBM: http://www-03.ibm.com/easy/page/6.

Due to this fact, lots of research has been carried out recently to test the accessibility of the interfaces, sometimes forgetting that this criterion (the provision of online material for everyone) is not enough to guarantee the ergonomics of the interfaces.

The first recommendations of WCAG 1.0 (Web Content Accessibility Guidelines), published in 1999, overlap with the recommendations suggested for accessibility: HTML language for web pages and CSS language for the style sheets of web pages.

As far as accessibility is concerned, 14 recommendations have been suggested in order to facilitate the provision of access for everyone[9]:

1) provide equivalent texts with visual or audio content;
2) do not rely on colors only;
3) use markers and style sheets correctly;
4) clarify the use of any natural languages;
5) create tables which can be dramatically transformed;
6) ensure that the pages which use new technology remain accessible and are displayed correctly;
7) ensure that the user controls content changes;
8) ensure the direct access to overlapping interfaces;
9) create a design which is independent of the hardware;
10) use intermediary solutions for older users of the interfaces;
11) follow the guidelines and technology of W3C technology;
12) provide contextual and directional information;
13) provide clear navigation mechanisms;
14) ensure that the documents are clear and simple.

The second version WCAG 2.0[10], which was released in 2006, is more likely to be subject to modifications. For the moment there are four key principles, each possessing several sub-criteria.

Principle 1: **perception**

1) provide technical solutions for all non-textual context;
2) provide synchronized solutions with multimedia applications;
3) ensure that the structure and information in the interface can be separated from the presentation;
4) make it easy to distinguish between information in the foreground and in the background.

---

9 WACG 1.0: http://www.w3.org/TR/WAI-WEBCONTENT/wai-pageauth.html.
10 WACG 2.0 : http://www.w3.org/WAI/WCAG20/quickref/.

Principle 2: **operability**

1) all functions can be accessed using the keyboard;
2) users control reading time or interaction time;
3) avoid illogical cutting of content;
4) provide mechanisms to help the users find content, to manipulate it and to manage it;
5) help the users to easily correct their mistakes.

Principle 3: **content and understandable controls**

1) readable and understandable texts;
2) placement and functionalities of predicted content.

Principle 4: **robustness**

1) compatibility of support with all users including the technology associated with help mechanisms;
2) provide content which can be easily accessed or provide accessible alternatives.

On the international level, the standards relating to accessibility published in 2003 under the international standard number ISO/TS 16071 have kept seven of the criteria which are very similar to the guidelines from WCAG 1.0:

1) provide the same solutions for visual and audio content;
2) do not exclusively use and rely on colors;
3) use markers and style sheets in an appropriate fashion;
4) clarify the use of any natural languages;
5) create tables which can be dramatically transformed;
6) ensure that the pages which contain new technologies can also be dramatically transformed;
7) ensure that the user is aware of any changes made to the content so that they are still in control of the content and not the other way around. A revised version of this standard number is currently under discussion under the standard number ISO/DIS 9241-171, which is being integrated into the 9241 series and which contains a lot of detailed information.

With the help of the Agency for the Development of Electronic Administration (Agence pour le développement de l'administration électronique or ADAE), which today is known as the General Service for the Modernization of the State (Direction générale de la modernisation de l'Etat or DGME), a reference system for the access to Internet services belonging to the French administration was created in 2004[11]

---

11 ADAE: http://www.adele.gouv.fr/spip/article.php3?id_article=246.

and a graphical and ergonomic charter for public online-procedures was created in 2006[12]. By relying on the association BrailleNet[13], the principles of accessibility were re-published in 2005 with 92 criteria that were grouped together into 13 smaller sections:

1) compulsory elements,
2) accessible content,
3) structure of the information,
4) presentation of the information,
5) help with navigation,
6) graphic elements,
7) links,
8) multimedia,
9) colors,
10) tables,
11) forms,
12) frameworks,
13) scripts.

15.2.3. *Publishers and ergonomic recommendations*

The introduction of library databases on the Internet is not a new concept. In the 1980s it was possible to carry out research on such databases using the ARPANET network. However, it has only been very recently that publishers have become interested in design guides for library interfaces on the Web so that the users themselves become familiar with using such sites. For example, the publisher Elsevier has integrated human-machine interfaces within its team of specialists. A guide aimed at maximizing the usability of library websites was published in 2004 [JAS 04]. In this guide eight criteria are stated:

1) *coherence*: use coherent design elements, provide a help tool at the top of each page;
2) *orientation and navigation*: use a navigation tool, directly link the library's web page to the home page of the institution;
3) *hyperlinks*: deal with hyperlinks according to agreements;
4) *page layout*: try to adjust the pages to avoid scroll bars;
5) *graphics and aesthetics*: do not use a lot of colors or graphics;
6) *flexibility and efficiency of use*: organize the information in several ways depending on the users' tasks and how often they use the site by minimizing the

---

12 DGME: http://synergies.modernisation.gouv.fr/rubrique.php3?id_rubrique=202.
13 Guide AccessiWeb: http://www.accessiWeb.org/fr/guide_accessiWeb/index.html.

number of clicks and give explanations to help the users choose the correct resources;

7) *make the system adequate for the user's activities*: clearly explain the research services without using jargon and ensure that the site functions properly;

8) *accessibility*: follow the recommendations from the W3Consortium/WAI[14].

On the European level the Minerva project, which is a ministerial network for the development of the digitization process, has a focus group which is involved in resolving issues related to quality, accessibility and usability. The guide "*Principles of quality in cultural Internet web sites*", which was published in 2003, states 10 principles[15]:

1) *identifiable*: the identity of the site is clearly defined as well as its objectives and the users should be able to clearly identify which organization is responsible for the site;

2) *relevant*: the content must be selected in relation to user needs, the resources and documentation must be accurate, navigation within the resources must also be made easy for the users;

3) *maintenance*: regular updates, renovations every six months, availability of the maintenance service;

4) *accessibility*: conforming to the W3C/WAI recommendations and other ways of accessing such sites;

5) *user-centered*: involve the users in all steps of the project and allow for user contribution;

6) *reactive*: the users can contact the site and receive the appropriate responses, the information is shared and discussed;

7) *multilingual*: content available in several languages;

8) *interoperable*: designing the site in accordance with appropriate standards;

9) *respectful of the law*: terms and conditions of access in relation to accessing the content;

10) *perennial*: make the content available online on a long-term basis.

At the end of this first section of the report it can be pointed out that when it comes to developing such sites and interfaces it is quite difficult to do so by only following the seven principles which form the standard number ISO 9241-110 or the other seven principles from ISO 9241-12, which is currently being revised. Each publisher, web designer and researcher works within a specific context in competition with one another and is obliged to offer interesting guidelines that must be taken into consideration when it comes to designing user interfaces. What, now, are the principles emerging from the current uses of digital libraries which are being investigated by researchers?

---

14 *Web Accessibility Initiative*: http://www.w3.org/WAI/.
15 Minerva: http://www.minervaeurope.org/publications/qualitycommentary_fr.htm.

15.3. Study of the uses of digital libraries

15.3.1. *Libraries and privileged relationships with the users*

Libraries, in competition with the Internet, have recently been forced to adapt what they have on offer to sustain their role as a provider of specialized information. Libraries have also showed a willingness to improve their services for their users [AND 04] by systematically evaluating user behavior. Worried about the needs of users and being in direct contact with them, the libraries have developed questionnaires which evaluate user satisfaction with the library services which are either left at the entrance of the library or distributed to the users by a specialist documentalist. These questionnaires also make it possible for the library to find out what the different needs of the online users are [CHA 00, REN 06].

References that focus more on the usability of human-machine interfaces have been available since 1994, during the first conference on digital libraries, which shows that some attention has been given to this type of interface [KLI 94]. However, during the Joint Conference on Digital Libraries A. Bradford and G. Buchanan insisted on the necessity, even the urgency, of ensuring that digital libraries are usable, useful and used. The authors highlight the fact that digital libraries have great potential in helping people organize, access and share information, but that the problem for the user is that it takes a certain period of time to become accustomed with how these libraries function. The majority of libraries do not provide enough appropriate advice to help the users gain a quick understanding of how the online libraries function.

The survey, which was carried out by 71 documentalists and led by Denise Troll Covey, highlights the real practices used in the evaluation of libraries [COV 02]. The good practice guide includes methods such as carrying out evaluations through the use of questionnaires, focus groups, evaluation protocols and analyzing user logs. A guide, which was published more recently by the National Science Digital Library, supports these evaluation methods and specifies the basics for the usability of interfaces [REE 05]. In addition, these two guides are quite similar to the standard number ISO/TR 16902 from 2002 (which will become ISO 9241-230) which details the methods of usability for user-centered design.

In October 2004 a special edition of the International Journal on Digital Libraries was devoted to the usability problems faced by digital libraries [BLA 04]. This special edition encouraged the digital library community to go further in relation to the research and application of the usability methods.

If traditional research has dealt with how to find information, provide it to the user and preserve the information, certain authors have already shown that new problems may arise, such as the tension that is created between the system's architects (designers, programmers) and the providers of information (authors, bookshops) [THE 00]. Other studies have dealt with the uses of systems in different workspaces, highlighting that shared spaces (supported by information professionals) can improve the suitability and circulation of new technology [ADA 05]. The social element in the design of the interface and its development can also have an impact on how well the resources are accepted by the users [ADA 04]. Finally, the role of users within certain workspaces is also being re-defined: students can go to the library several times a day to write emails, make photocopies or use the Web for personal or recreational reasons. They can also purchase items online, check their orders for interlibrary loans, borrow a book, talk to a friend online or read magazines or newspapers [DEM 05]. The role of the documentalist is also changing: more than 274 tasks were mentioned by the 46 American libraries that took part in the survey; the tasks are grouped together in six larger categories [CHO 06]:

1) management,
2) technology,
3) processing,
4) digital libraries,
5) collections and resources,
6) other.

15.3.2. *Getting lost in digital library interfaces*

Getting lost in digital library interfaces remains a key issue [THE 00]. The users of the interfaces create a mental representation of the system and use this mental representation when interacting with the actual digital libraries. When the users already have a certain representation of the structure's system and this representation does not work with the system that they are actually using, the users (both beginners and experts) then consider themselves to be lost. The degree of being lost varies according to the portals tested. The users believe that the portals which were used in the test did not provide them with enough information to help them understand the conceptual model used by the digital libraries.

A lot of the problems come from the fact that the users navigate on the site quite randomly which does not enable them to meet their aims [BLA 01]. For example, researchers tested six different libraries with five users. Some of the users spent more than six minutes trying to download an article from one of the libraries and then finally gave up saying that they had tried everything, but that they could not

download the article. The users also have a tendency to confuse search functions and browse functions. According to the authors, five key points should be considered when designing interfaces:

1)   *familiarity*: the users have to be able to quickly understand the basic characteristics of the digital library, the content and the structure in order to carry on with their research on the library websites;

2)   *deadlines*: numerous requests are unsuccessful for the users. The absence of a relevant response can sometimes lead to a better understanding of the database's structure, however, if the users feel submerged by the results they will have the feeling of not being able to obtain the relevant documents because it is difficult to distinguish between the documents;

3)   *being able to distinguish between different pieces of information:* it must be ensured that it is possible for users to be able to distinguish between different pieces of information made available to them;

4)   *serendipity*: the discovery of unexpected results can give the users a sense of progress. Serendipity depends on the user's capabilities of identifying information, i.e. the users can distinguish between different pieces of information;

5)   *changes in context*: the transition events, which imply that an agent (the user or the system) changes the context, can create problems during the interaction. Measures must be taken in order to insure that the agents maintain a common base and understand the consequences of transitions.

15.3.3. *The use of online catalogs and databases*

Numerous studies have looked at how users use the interfaces of online catalogs. If the users are able to search for and find information amongst the collections of four digital libraries such as the IEE-CS, ACM, NDLTD and NCSTRL, it must be pointed out that the search time changes depending on the digital library. However, overall user satisfaction as far as websites are concerned is linked to the site's efficiency, i.e. the site in which the user makes the least number of errors rather than the site which has the shortest search time when searching for online documents [KEN 99].

Authors from four universities in New Zealand have found 98 usability problems in relation to the users of online catalogs. These problems have been grouped together into six categories [WHI 06]:

1)   problems related to the organization of too much information;
2)   lack of coherence which reduces the ease with which users can use the catalogs;
3)   the absence of certain research functions;
4)   the terminology that is used;

5)   lack of feedback;
6)   irrelevant help tools.

The authors suggest that the interface should be simplified, the terminology should be improved, and the catalog should be modernized to incorporate links with resources on the Internet. The classification below is more or less similar to the 28 usability problems found by [HAR 04] who groups the problems into seven wider categories:

1)   feedback,
2)   coherence,
3)   graphic design and organization,
4)   terminology,
5)   navigation,
6)   system model,
7)   research and browse functions.

Some of the characteristics of business sites (Questia and Blackboard) can also be used to help design the websites of university libraries (the University of Arizona, UA and the University of California, CSULB). The overall productivity of the site, navigation on the site and the ease of finding specific information are much better for the website of the University of Arizona's library and the business site Questia. In total, the users are able to carry out 67% of all tasks requested. The users can carry out 78% of all tasks on the University of Arizona's website and 73% of all tasks on the Questia website, whereas on the University of California's website, 59% of all tasks can be carried out and this figure reaches 52% for the Blackboard website. The authors go into detail about the problems that were encountered on each site and describe how the young people using these sites do not assimilate the library catalogs as full text resources. The library is therefore more of a physical link where people come to carry out research, make photocopies and look for documents.

The recent report prepared by the Library of Congress shows that students of all ages avoid using library catalogs and prefer to use new tools. If the library catalogs are currently being under-used they will, however, be essential in the years to come as a tool to search through library collections [CAL 06]. They will also become essential because of the fact that there is going to be a rapid increase in the number of digitization projects [KAH 06]. The library catalog therefore needs to be re-worked to make it more visual and easier to explore, for example with interactive visual cards [PAP 05]. The default options that are displayed also need to be checked so that they are appropriate for the majority of users because 33% of requests are never modified by the users [JON 98]. Nevertheless, other studies show that the catalog is normally the first point of contact for students (51.9%), followed

by web pages (28.6%) or printed books and articles in the library (15.8%) and these figures rise sharply if the students are studying information sciences [LIU 06].

The fact that students are more at ease with online research than with using a piece of documentation software suggests that there is a conflict between the two representations [DES 06]: on one hand, there are the library users who have to deal with the restrictions of the library and its opening times, on the other hand, there are the Internet users who have a sense of belonging to an online community that can provide them with useful information such as reading notes and the opinions of other members of the online community (which is explained in everyday language); this information is considered essential in refining the pre-selection of material.

The libraries of the University of Ankara studied the use of online databases to check if subscriptions to electronic resources are useful for the researchers [ATI 06]. First of all, are researchers aware of the existence of the resources? What is the level of usage of digital resources amongst the university staff and what are their preferred library databases? Almost 2,000 people were surveyed in 2002 using a questionnaire; the results show that:

a) a large majority of the respondents know that the electronic resources exist (86.5%). Awareness of such resources varies according to the hierarchical level within the university: associate professors (93.3%), assistant professors (90.8%), professors (89%), research assistants (88.6%), lecturers (84.7%), specialists (83.5%) and finally teachers (31.7%). The authors assume that there is a better awareness amongst the associate and assistant professors because they hope to receive a promotion through their own original research and publications, which is not the case for the teachers;

b) amongst the respondents who knew that the electronic resources existed, a quarter of them said that they did not know anything about the databases, nearly half of the respondents said they knew something about certain databases and almost a third of respondents knew a lot about the databases;

c) nevertheless, 52 % use the databases occasionally, 27.5% often and 20.5% not at all;

d) the results support the idea that the use of the electronic databases is influenced by the level of knowledge that the users have of them and that the librarians should give more importance to teaching people how to use the databases.

Norris [NOR 05] studied the website at the library of the University of Wollongong in Australia using three tests:

1) an examination of the interface by experts;
2) the observations of seven tasks carried out by 25 participants;
3) post-questionnaire on the perception of the site.

In terms of examining the interface three lists of criteria were used: the list from the National Cancer Institute [LEA 06], Gerry Gaffney's[16] check-list from 1998 and Raward's[17] check-list from 2001 [RAW 01]. The first two lists made it possible to find 86% and 85% of the interface's problems respectively. However, Raward's check-list only found 80% of the problems linked to usability. Even if these two lists perform quite well they still remain rather limited when it comes to pointing out problems that are specific to libraries, for example the variety of search possibilities which must be examined in more depth. Raward's check-list seems to deal more specifically with the library's requirements, however, it does not perform as well when it comes to problems related to the usability of the interface. The authors believe that in order to completely evaluate the usability of a library's website it is necessary to consider both the general usability problems as well as the problems that are specific to libraries.

15.3.4. *Listening to the researchers' needs*

Whilst some studies rely on a questionnaire and a series of individual interviews from a sample of researchers from Faculties of Humanities, other studies investigate the practices of researching online documents and the expectations of the public, as is the case at the Denis Diderot library which groups together three distinct libraries from Faculties of Humanities [CHA 06]. The study shows an increasing use of digital libraries: 88.7% of researchers from the Faculty of Humanities said that they regularly search for online documentation on the Internet either at home or in the library. There is a large variety of resources that can be consulted online: more than 84 references were quoted by the 17 researchers that were questioned. The majority of the researchers stated that accessing the documents was rather complicated or difficult. A lot of the documents are referenced, but the users have problems getting exactly what they are looking for. Several wishes were expressed by the users and it is possible for the digital libraries to respond to the requests that have been expressed by users. The main requests of the users were as follows: that the online libraries provide the latest information to the users, that the libraries provide collections of works on a European level, that the users do not have to be within the library in order to access the library network. The users also requested that a guide be created (and which can be found at the entrance of the library) for the Denis Diderot library in order to have a better understanding of the layout of the library and to understand what services and collections the library has to offer. The final suggestion that was made by the users of the libraries was to break the anonymity with the documentalists because the researchers have problems in identifying the different people and their roles within the library (for example finding the correct

16 Gaffney: http://www.infodesign.com.au/usabilityresources/evaluation/Webevaluation.asp.
17 Raward: http://alia.org.au/publishing/aarl/32.2/full.text/raward.html.

person in the organizational charts, finding out the roles, names, telephone numbers and email addresses of these people and even being introduced to them), as well as for the need for documentalists on site because the researchers can spend up to an entire day in the library.

15.3.5. *User-centered focus groups within libraries*

In addition to the studies which have been carried out through the use of questionnaires given to users, some libraries have created user-centered focus groups. For example, Western Washington University Library[18] have gone beyond analyzing user impressions and satisfaction from forms filled out on the Web, instead in 2005 they introduced a one and a half hour long focus group, made up of eight participants, in order to understand the needs of the users in relation to the digital library's website. The problems mentioned by the users concern online help in relation to the websites, the organization of the site, navigating through the site, terminology, the representation of the database content, the sites' performance, the relevance of search results, the services that are on offer, access to the site and training on how to use the site.

The University of Washington library[19] also introduced usability studies in relation to their library's interface. In 2004, 82% of those people surveyed accessed the online catalog, 51% accessed the online database and 65% stated that they could easily, or even very easily, find what they were looking for.

Other universities prefer to take part in quality studies known as LibQual+[20]. The results of these quality studies can then be integrated into the online libraries.

15.3.6. *Suggested recommendations for improving digital libraries*

From early on, research has shown certain recommendations that could be used for the improvement of digital libraries. Certain authors [KLI 94] have relied on the research carried out by Nielsen [NIF 93] in order to come up with 11 ergonomic recommendations:

1) know the user,
2) analyze the competition,
3) establish a minimum level of usability,
4) offer several different designs for the interface,

---

18 Western Washington University Library: http://www.library.wwu.edu/udt/.
19 University of Washington DC Libraries: http://www.lib.washington.edu/usability/.
20 LibQual+: http://www.libqual.org/.

5) user participation,
6) coherence of the interface resulting from a high level of co-ordination,
7) choice of appropriate guidelines for solutions,
8) creation of user handbooks,
9) user tests,
10) iterative design,
11) gather all of the usability results together for future designs.

For other authors three main factors could improve the design of digital libraries [THO 04]:

1) *the characteristics of the interface*: clear and comprehensive terminology with no jargon. The screen should be well organized and navigation within the interface should be logical and uncomplicated;

2) *organization*: provide relevant databases in order to answer user requirements, it should be possible to access the system from certain designated computers and the system should be easy to understand;

3) *individual differences*: reinforce the efficiency of the interface by providing training, and using other pieces of software to improve the flexible design of the interfaces in relation to the user's knowledge of the system, i.e. the interface will become more technically advanced depending on whether or not the users know how to use the system.

Bevan and Klinca[21] rely on the National Cancer Institute guide from 2003 [LEA 06] when it comes to particular supplementary recommendations for digital libraries. Three elements characterize digital libraries:

1) they contain a large quantity of information;
2) it is difficult to determine what the users want;
3) needs refer to both browsing and searching for information.

Design and evaluation: in order to create an interface that works well it is necessary to understand the who, the why, the context and also to take inspiration from the services that have interfaces which are already working successfully. It must be ensured that the users are able to concentrate on their task in hand rather than focusing on how the digital library functions.

Links: links to related information do not have to be included within the documents, meaning that users do not have to go through lots of documents in order to find links to related information which is time consuming. However, the users will have to carry out other searches and must wait for a response which can be time-consuming.

---

21 Bevan and Klincla: http://www.usabilitynet.org/guidelines/digital_libraries.html.

Content: the documents must be easy to read or run. Users should not be submerged in lots of texts, the texts should be relevant and easy to understand. The headers of documents are important for identifying the document's content and the footnotes make it possible to identify the origin of the document, the author of the document, contacts, the date of creation of the document and copyright. The system must resemble the real world as much as possible in such a way that the digital library can be integrated within the everyday workplace.

The search for information: information must be provided on all the digital libraries that are currently available, and an idea of the content each of these libraries holds should also be given. It should be possible to explore a library in order to understand how a collection of works covers a particular theme. There should also be some indication to show the users whenever they enter or leave the research area of a particular library. The users do not want to receive all of the results available, they prefer a manageable number so that they can distinguish the content of the results. Users tend to give up looking for a relevant document unless they find what they are looking for within a relatively short period of time. The users want to have the latest information, interesting documents and documents which shock and surprise them. It should be possible for the users to save, modify and go back to previous search requests in order to use them again. Different search forms should be made available to the users depending on their experience. Jargon which is used by the documentalists should be avoided as much as possible; abbreviations and acronyms should only be used for the experienced user.

From experience, libraries are also developing guidelines in order to further improve and develop their web sites. For example, Australian libraries[22] suggest only five recommendations:

1) *information to be displayed somewhere on the site*: a) aims and objectives, b) contacts, c) services, d) latest news;
2) *information to be displayed on each web page*: title, name of the library, date of the latest modification, link to the home page;
3) *information on certain key pages*: contact for feedback, URL of the website in case users are printing documents from the site;
4) *metadata*: title, description, keywords and author so that this information can be found through the use of search engines;
5) *accessible* to everyone.

Some general principles of good practice which are to be considered in the creation of such sites and interfaces, which were taken from university consortiums [WHI 04], are listed below:

---

22 Australian Libraries: http://www.nla.gov.au/libraries/resource/guidelines.html.

1) keep navigation simple and ensure that users can go back to the library's home page;

2) provide various access routes to the site in order to satisfy the different ways in which people carry out research;

3) avoid using too much text and avoid confusing the users;

4) avoid using jargon and ambiguous terms;

5) do not make any assumptions about user behavior. For example, dyslexic students can use the sites much better than university academic staff.

15.3.7. *Recommendations based on user opinions*

The criteria used in the evaluation of digital libraries can also be suggested by the users of the libraries themselves.

In 2003 and 2004, 48 students developed and justified the use of a set of criteria which, for them, seemed important in the evaluation of digital libraries [XIE 06]. The usability and quality of the collection of works available were considered as essential criteria by the students. These criteria were followed, in order of importance, by the quality of the service, efficiency and performance level of the system, and finally obtaining user opinion.

15.4. Conclusion

What conclusion can be drawn from this presentation of the ergonomic standards of digital libraries and their associated uses? This report has highlighted the difficulties of using these libraries, as well as highlighting the difficulties linked to the complex interfaces in which the users do not have an overall view of the databases that can be explored, they also feel lost and drowned in technical vocabulary and have the impression of not really finding what they are looking for. What can be recommended to the interface designers of digital libraries?

One fact becomes clear straight away: there are too many standards and guidelines for human-machine interfaces. These standards and guidelines, which are continually evolving, vary according to different researchers, computer engineers and publishers. On one hand, more precise guidelines have been suggested in order to help with the construction of the interfaces, on the other hand more specific recommendations have been put forward in order to solve issues related to a particular page or a given problem. The standards are more generalized and the guidelines are more precise but at the same time they are not as apt at dealing with specific tasks.

However, if libraries are literally collapsing under the strain of the number of standards they have to respect in relation to indexing, metadata, intellectual property, the interoperability of resources, programming languages, access controls, accessibility, etc. [McC 01], they have integrated the concepts of usability from a very early stage [SHI 03]. Research shows that if potential users of these sites are aware of what resources are available to them, they will still be unable to use the resources. Several factors can influence the reasons behind people's use of and reaction to technology [THO 02]: the perceived ease in using such technology in general and then the perceived use of this technology by the users themselves. It is clear that the interfaces of these digital libraries are not as easy to use as might have been initially thought, due to the richness and complexity of the information available. Why are the interfaces of digital libraries not as easy to use as it was imagined they would be?

From the literature that has been analyzed it can be stated that libraries focus on four types of people: information specialists (documentalists/archivists/publishers), information technology specialists (designers/developers/administrators), library users and specialists in the field of usability (ergonomists/communication science), i.e. this means that there are four different views as far as interfaces are concerned. There is the everyday view for the people who work and spend all their time in the library, or the occasional view for the people who use the library. Then there is the more exterior view, or view from the outside for those people who study and research interfaces, and finally there is the more technical view such as the 5S structure[23] for designers, as suggested by [GON 04]. Decisions relating to the design of interfaces are usually influenced by the design and development team, and by those who provide the information. The decisions which are made all directly impact the users of the websites [XIE 02].

Specialists in the field of usability are not always present in libraries. Unless they are involved in the project from the beginning (from the design phase), they will probably not have much influence on the technical aspect of the project. The documentalists, who are involved from the beginning, bring their knowledge of user-centered designs to the project and are encouraged to study the interfaces and can replace the ergonomists working on the project. However, the documentalists do not have much influence over the designers of the interfaces nor over the interfaces themselves. The users of the websites seem to be asked their opinion after the design of the website and the interface rather than being involved in the project from an earlier stage. Therefore, it is important to agree on a multidisciplinary view of the projects being worked upon.

---

23 5S: "*Streams*, *Structures*, *Spaces*, *Scenarios*, *Societies*".

New ergonomic criteria which are specific to the interfaces of digital libraries are emerging. These criteria are additional to the 9241 standards and the accessibility standards which are essential for libraries to take on board. The new criteria focus on:

1) the representation of the databases, their content and their transitions;
2) the quality of the collection of works available;
3) the efficiency of the system;
4) navigation, search, browse and full-text reading functions;
5) coherence, given the variety of resources available;
6) personalization by domain;
7) modernizing and updating the library's online catalog by including links to other resources on the Internet within the catalog;
8) the services which are common to both the physical library and the digital library;
9) online aids;
10) training.

More work needs to be done in the (re)definition of the ergonomic standards that are specific to digital libraries to allow a closer look at all of the concepts related to digital libraries, just as Bevan is currently trying to do for the ergonomic standards [BEV 01, BEV 05, BEV 06]. The portals of libraries have the potential to become very rich in content and could provide a set of services which would involve the collaboration of the library and the interface [DEE 06] and would make all the difference in the project i2010: Digital Libraries Initiative or in open access projects which would be re-examined by investigating both usability and accessibility.

There has still been no general agreement on the usability of online libraries and studies have highlighted that these online libraries are under-used. However, the time saved resulting from the fact that users are able to access online libraries from home or from work, coupled with the relevance of the information obtained through research, are reassuring, at least in terms of the usefulness of the online libraries.

15.5. Bibliography

[ADA 04] ADAMS A., BLANDFORD A., "The unseen and unacceptable face of digital libraries", *International Journal on Digital Libraries*, no. 4, p. 71-81, 2004.

[ADA 05] ADAMS A., BLANDFORD A., LUNT P., "Social empowerment and exclusion: a case study on digital libraries", *ACM Transactions on Computer-Human Interaction*, vol. 12, no. 2, p. 174-200, June 2005.

[AFN 03] AFNOR, *Ergonomie de l'informatique. Aspects logiciels, matériels et environnementaux*, Collection of Computing Standards, Paris, 2003.

[AND 04] ANDREWS J., LAW D., *Digital Libraries: Policy, Planning and Practice*, Ashgate Publishing, London, 2004.

[ATI 06] ATILGAN D., BAYRAM O.G., "An Evaluation of Faculty Use of the Digital Library at Ankara University, Turkey", *The Journal of Academic Librarianship*, vol. 32, no. 1, p. 86-93, 2006.

[BAE 87] BAECKER R.M., BUXTON W.A.S., "The star, the Lisa and the Macintosh", in *Readings in Human-Computer Interaction, A Multidisciplinary Approach*, p. 649-652, Morgan Kaufmann Publishers, New York, 1987.

[BEV 01] BEVAN N., "International standards for HCI and usability", *International Journal of Human-Computer Studies*, vol. 55, no. 4, p. 533-552, 2001.

[BEV 05] BEVAN N., "Guidelines and Standards for Web Usability", *Proceedings of Human Computer International 2005*, CD, 2005.

[BEV 06] BEVAN N., "International Standards for HCI", based on a chapter in the *Encyclopedia of Human Computer Interaction*, Idea Group Publishing, 2006, http://www.usabilitynet.org/papers/International_standards_for_HCI.pdf.

[BLA 01] BLANDFORD A., STELMASZEWSKA H., BRYAN-KINNS N., "Use of Multiple Digital Libraries: A Case Study", *JCDL'01*, p. 179-188, Roanoke, Virginia, 24th–28th June 2001.

[BLA 02] BLANDFORD A., BUCHANAN G., "Usability for Digital Libraries", *Proceedings of the 2d ACM/IEEE-CS Joint Conference on Digital Libraries JCDL'02*, p. 424, Portland, Oregon, 13th–17th July 2002.

[BLA 04] BLANDFORD A., KEITH S., CONNELL I., EDWARDS H., "Analytical usability evaluation for digital libraries: a case study", *Proceedings of the 4th ACM/IEEE-CS Joint Conference on Digital Libraries JCDL '04*, p. 27-36, 2004.

[CAL 06] CALHOUN K., The Changing Nature of the Catalog and its Integration with Other Discovery Tools, Final report for Library of Congress, 17th March 2006, http://www.loc.gov/catdir/calhoun-report-final.pdf.

[CAN 82] CANFIELD SMITH D., IRBY C., KIMBALL R., VERPLANK B., "Designing the Star User Interface", reprinted from *Byte*, no. 4, p. 242-282, 1982, http://www.guidebook gallery.org/articles/designingthestaruserinterface.

[CHA 00] CHADWICK J., FALK J.H., O'RYAN B., "Assessing Institutional Web Sites. Summary of Report", in *Collections, Content and the Web*, *Council On Library Resources*, p. 64-78, Council on Library and Information Resources, Washington D.C., 2000.

[CHA 06] CHARRA G., Practices used in documentary research and the expectations of researchers within the Faculties of Letters and Humanities. Study of the Denis Diderot library, Dissertation from the ENSSIB, January 2006, http://www.enssib.fr/bibliotheque/documents/dcb/charra.pdf.

[CHO 06] CHOI Y., RASMUSSEN E., "What do Digital Librarians do ?", *Proceedings of the 6th ACM/IEEE-CS Joint Conference on Digital Libraries 2006*, p. 187-188, Chapel Hill, NC, USA, 11th–15th June 2006.

[COU 90] COUTAZ J., *Human-Computer Interfaces, Design and Creation*, Dunod, Paris, 1990.

[COV 02] COVEY D., Usage and Usability Assessment: Library Practices and Concerns, Digital Library Federation and Council on Library and Information Resources, Washington D.C., 2002, http://www.clir.org/PUBS/reports/pub105/pub105.pdf.

[DEE 06] DEE C., ALLEN M., "A Survey of the Usability of Digital Reference Services on Academic Health Science Library Web Sites", *The Journal of Academic Librarianship*, vol. 32, no. 1, p. 69-78, 2006.

[DEM 03] DEMPSEY L., "The Recombinant Library: Portals and People", *Journal of Library Administration*, vol. 39, no. 4, p. 103-136, 2003.

[DEM 05] DEMAS S., "From the Ashes of Alexandria: What's Happening in the College Library?", *Library as Place: Rethinking Roles, Rethinking Space*, CLIR Publication, no. 129, p. 25-40, 2005, http://www.clir.org/pubs/reports/pub129/pub129.pdf.

[DES 06] DESPRES-LONNET M., COURTECUISSE J.-F., "Les étudiants et la documentation électronique", *BBF*, vol 51, no. 2, p. 33-41, Paris, 2006.

[GON 04] GONÇALVES M.A., FOX E.A., WATSON L.T., KIPPE N.A., "Streams, Structures, Spaces, Scenarios, Societies (5S): A formal model for digital libraries", *ACM Transactions on Information Systems*, vol. 22, no. 2, p. 270-312, April 2004.

[HAR 04] HARTSON H.R., SHIVAKUMAR P., PÉREZ-QUIÑONES M.A., "Usability inspection of digital libraries: a case study", *International Journal on Digital Libraries*, vol. 4, no. 2, p. 108-123, October 2004.

[HOR 06] KASPER HORNBÆK K., "Current practice in measuring usability: Challenges to usability studies and research", *International Journal of Human-Computer Studies*, vol. 64, no. 2, p. 79-102, February 2006.

[JAS 04] JASEK C., "How to Design Library Web Sites to Maximize Usability", *Library Connect* in collaboration with Elsevier User Centered Design Group, 2004, http://www.elsevier.com/framework_librarians/LibraryConnect/lcpamphlet5.pdf.

[JON 98] JONES S., CUNNINGHAA S.J., MCNAB R., "Usage analysis of a digital library", *Proceedings of the Third ACM Conference on Digital Libraries*, p. 293-294, Pittsburgh, Pennsylvania, 23rd–26th June 1998.

[KEN 99] KENGERI R., SEALS C.D., HARLEY H.D., REDDY H.P., FOX E.A., "Usability study of digital libraries: ACM, IEEE-CS, NCSTRL, NDLTD", *International Journal on Digital Libraries*, no. 2, p. 157-169, 1999.

[KLI 94] KLING R., ELLIOTT M., "Digital library Design for Usability", *DL'94*, *First Annual Conference on the Theory and Practice of Digital Libraries*, College Station, TX, USA, 19th–21st June 1994, http://www.jcdl.info/archived-conf-sites/dl94/paper/kling.html.

[LEA 06] LEAVITT M.O., SHNEIDERMAN B., Research-Based Web Design & Usability Guidelines, PDF file from the US Gov. Printing Office, 2006, http://www.usability.gov/.

[LIU 06] LIU Z., "Print vs. electronic resources: A study of user perceptions, preferences, and use", *Information Processing and Management*, vol. 42, p. 583-592, 2006.

[MCC 01] McCray A.T., Gallagher M.E, "Principles for digital library developpement", *Communication of the ACM*, vol. 44, no. 5, p. 49-54, May 2001.

[NIE 93] Nielsen J., *Usability Engineering*, Morgan Kaufmann Publishers, New York, 1993.

[NOR 05] Norris A. Freeman M., "Improving library Website usability: a user focus", *neXt 2005: ALIA National Library and Information Technicians Conference*, 2005, http://e-prints.alia.org.au/archive/00000062/01/NorrisImprovingLibraryWebsiteUsability.pdf.

[ODA 93] O'Day V.L., Jeffries R., "Orienteering in an Information Landscape: How Information Seekers Get From Here to There", *Proceedings of InterCHI '93*, p. 438-445, 24th–29th April 1993.

[PAP 05] Papy F., *Les bibliothèques numériques*, Hermes, Paris, 2005.

[POM 06] Pomerantz J., Wildemuth B.M., Yang S., Fox E.A., "Digital library curriculum: Curriculum development for digital libraries", *Proceedings of the 6th ACM/IEEE-CS Joint Conference on Digital Libraries JCDL '06*, p. 175-184, 2006.

[RAS 06] Raza M.M., Upadhyay A.K., "Usage of E-journals by researchers in Aligarh Muslim University: A study", *The International Information & Library Review*, 2006, no. 38, p. 170-179.

[RAW 01] Raward R., "Academic Library Website Design Principles: Development of a Checklist", *Australian Academic & Research Libraries*, vol. 32, no. 2, June 2001, http://alia.org.au/publishing/aarl/32.2/raward.html.

[REE 05] Reeves T.C., Apedoe X., Woo Y.H., "Evaluating Digital Libraries: A UserFriendly Guide", *University Corporation for Atmospheric Research*, July 2005, http://www.dpc.ucar.edu/projects/evalbook/EvaluatingDigitalLibraries.pdf.

[REN 06] Renoult D., "Enquêtes de publics dans les bibliothèques universitaires. Où en sommes-nous ?", *BBF*, Paris, vol. 51, no. 2, p. 5-9, 2006.

[SCA 86] Scapin D.L., Ergonomic guide in the design of human-machine interfaces, technical report from the INRIA, RT77, October 1986, http://hal.inria.fr/view_by_stamp.php?label=INRIA-RRRT&langue=fr&action_todo=view&id=inria- 00070083&version=1.

[SHI 03] Shiri J., "Digital library research: current developments and trends", *Library Review*, vol. 52, no. 5, p. 198-202, 2003, http://eprints.rclis.org/archive/00001521/.

[THE 00] Theng Y.L., Modh-Nasir N., Thimbleby H., "Purpose and usability of digital libraries", *Proceedings of the Fifth ACM Conference on Digital Libraries*, p. 238-239, San Antonio, Texas, June 2000.

[THO 02] Thong J.Y.L., Hong W., Tam K.-Y., "Understanding user acceptance of digital libraries: what are the roles of interface characteristics, organizational context, and individual differences?", *International Journal of Human-Computer Studies*, no. 57, p. 215-242, 2002.

[THO 04] Thong J.Y.L., Weiyin H., Kar Y.T., "What leads to user acceptance of digital libraries?", *Communications of the ACM 2004*, vol 47, no. 11, p. 79-83, 2004.

[TRA 02] TRAVIS TIFFINI A., NORLIN E., "Testing the Competition: Usability of Commercial Information Sites Compared with Academic Library Web Sites", *College & Research Libraries*, p. 433-448, 2002.

[WHI 04] WHITE S., "Assessing library Web page usability: how benchmarking can help", *Library and Information Research*, vol. 28, no. 88, p. 47-52, Spring 2004, http://eprints.rclis.org/archive/00003487/01/88_white.pdf.

[WHI 06] WHITE H., TIM WRIGHT T., CHAWNER B., "Usability Evaluation of Library Online Catalogues", *Seventh Australasian User Interface Conference, AUIC2006*, 2006, http://crpit.com/confpapers/CRPITV50White.pdf.

[XIE 02] XIE H.I., WOLFRAM D., "State Digital Library Usability: Contributing Organizational Factors", *Journal of the American Society for Information Science and Technology*, vol. 53, no. 13, p. 1085-1097, 2002.

[XIE 06] XIE H.I., "Evaluation of digital libraries: criteria and problems from users' perspectives", *Library & Information Science Research*, vol. 28, no. 3, p. 433-452, 2006.

Chapter 16

A Document Information System Within the University: From the Project's Conception to its Installation

The research department of the University of Artois has set as its main target the aim of providing documentation and services which are increasingly adapted to the needs and expectations of the users of such resources. This goal lies within the framework of the university's traditional aims and objectives which are linked to the organization, conservation and provision of documentation. In trying to reach its goal, the university also integrated the changes which have affected the provision of documentation; changes which have been brought about by the rapid increase in the digitization of documentation.

In order to reach this objective, the research department of the university defined three key areas which form the basis of its development project:

1) to become more involved in the fundamental aims and objectives of the university (i.e. more involved in teaching, research etc.):

– the definition of a multi-level and multi-support acquisition policy,

– the creation of specialized stocks of documents in close co-operation with teachers and researchers,

– the development and evaluation of material that is available online;

Chapter written by Corinne LEBLOND.

2)   to make all of the library services widely available to the different types of library user:

– a general reflection on the welcoming and orientation of users, as well as analyzing the information that is available to users,

– the renewal and revision of training provided for users on the resources available to them at the university's library;

3)   to adapt the services, which are provided to the users, in relation to user needs and expectations:

– re-computerization of the library network,

– provision of cultural activities,

– evaluation of issues linked to the documentation and its associated uses,

– evaluation of user satisfaction,

– document engineering.

The document information system, which is at the heart of this policy of modernizing the services available to the general public, is one of the major factors behind both the revival of relationships with the different services within the university and the change in the services currently provided by the university library. The document information system at the University of Artois is made up of an information service and a training service that enable its users to carry out the best research possible in order to increase their knowledge of a particular topic. The document information service provides flexible physical and virtual resources which can be upgraded in order to be adapted to all sorts of policy changes that take place within the university.

Thanks to the integration of Web 2.0, and in particular RSS feed within the system, the user is able to create their own personal digital workplace by actively participating in the construction of the access interface which provides access to increasingly diverse digital resources and services. As a result of this, it is possible to combine the processes of individual and collective construction depending on the approach used by the user to research information.

This project is a continuation of the modernization of the services mentioned previously which sees the process of documentation as an ever-evolving process. One of the main factors behind the success of this project is the ability to integrate the project within the university's document information system.

Prior to the introduction of this new project, and in order to precisely define the context in which this project will evolve, we need to understand the recent history of the university and its document information system.

16.1. Where do the university and its document information system originate from? Conditions for use of such a system.

16.1.1. *Local context: the document information system within the university*

The University of Artois, which was once a campus of the University of Lille, became autonomous on 1st October 1992 as did the university's research department.

When the university opened its doors in October in 1992 it had three campuses, with each campus specializing in different areas of study:

– Arras: humanities,

– Béthune: economic and applied sciences, as well as an Institute of Technology,

– Lens: hard sciences.

The university then expanded with two other campuses:

– Douai: legal and political sciences,

– Liévin : sports science.

An Institute of Technology was also created at the Lens campus.

Today the university has 11,000 students and a library on each campus. The university also combined the libraries of the two Institutes of Technology.

Thanks to a policy on the centralization of documentation that was established when the university was created, the libraries of each campus gathered together all of the documents and resources related to teaching and research within the university. This meant that no departmental libraries or laboratories were created within the university.

The university's research department plays a key role in the university when it comes to:

– the composition and display of collections of works,

– the display of document resources produced by the university.

The research department also has the responsibility of maintaining the stocks of resources and services provided by the university for:

– the needs and expectations of all users,

– the different areas of study,

– training and research needs.

16.1.2. *The emergence and development of a regional online university*

At the same time as the evolution of the document information system at the University of Artois, another process, this time related to documentation services, was taking place, i.e. the implementation of a regional online university based on the digital workspace, known as the ESUP Portal.

The universities from the Nord-Pas-de-Calais region in France, as well as the National College of Arts and Textile Engineering (École Nationale Supérieure des Arts et Industries Textiles or ENSAIT), created a consortium in 2003 aimed at generalizing the use of a digital workspace within the region. This digital workspace can be adapted to the needs of both the students and staff at the universities. The other aim of the consortium was to ensure a coherent development of digital resources throughout the entire region.

In order to achieve these objectives the project was divided into several parts:

– part 1: the use of the digital workspace (ESUP-Portal), the users access information that is relevant to them; the information is grouped into various channels with each channel being devoted to the needs of a particular group of users;

– part 2: the creation of a free access computer room so that users can access the information they need as and when they wish;

– part 3: the development of new solutions which favor student mobility;

– part 4: the development of a regional portal on which the digital workspaces of the universities can place all of their documentation;

– part 5: the sharing of training methods and the simultaneous access of information by different users for optimal use of the digital workspaces.

In using this approach and in order to respond to the issues behind each of the five parts mentioned above, it is important for the research department of the university to re-affirm its role as the document information system of the university whatever the subject, level of the students and medium the students have used in order to access the information.

16.2. The implementation of the document information system

16.2.1. *The success of the modernization of documentation*

A spreadsheet can easily place the different stages of the computerization of our documentation service in chronological order. The document information service, which was the success of the modernization of our documentation service, enables us to detect the frequency and density of technological evolutions, which over time have shaped the services that we have on offer for the users of our document information system.

The first phase, which was successfully completed in 1996, involved the installation of the integrated library system which deals with all management tasks and library tasks.

The interface access, which enables users to access the online catalog, was made available to the general public in 1998 and has been available online since 2000.

The third phase in 2002 saw the integration of Web OPAC for the entire network.

The fourth and final phase consisted of integrating the University of Artois document information system within the university's general information system.

Since 1998, the University of Artois' research department has been working on a project to modernize the tools related to the management and control of document information. The research department introduced a CDROM network (which at that time was a vital support tool for online information), and also installed computers within the libraries, these computers were to be used specifically for researching information.

The invitation to bid which was launched for this project proved to be unsuccessful. As of September 2003 the university's research department re-launched its project by changing some of its ideas based on the experience it had from the project launched in 1998, and also by taking into consideration the evolution of products and user needs.

Even if the objectives of the project that was relaunched in 2003 remained the same as the objectives of the initial project that was launched in 1998, the overall structure of the system has changed. The document system incorporates a consultation tool that enables users to consult online resources independently of the medium or computer that the users are using, or even from where they are accessing the information. The system is the basis of a documentation portal which is linked to

document management software. If we want to develop and improve the number of online documents that are available, an access and research interface needs to be created to insure the maximum potential of the resources and the suitability of the material that is available online for the users.

In this respect, the document information system was designed to be integrated within the university's more general information system and also to give access to a set of:

– services:

    - access to a set of services and resources from only one interface, which incorporates the library and university's graphic charter,

    - access to personalized services: registered-user personal basket, search history, saved personal searches, access to reader file (such as current loans and reservations, etc.),

    - access to important information prepared by the research department: folders of documents which are grouped together either by theme or by area of study, as well as the latest document acquisitions;

– resources:

    - the catalog,

    - external online resources which have been chosen or acquired by the research department:

        - an index of Internet sites,

        - electronic databases or periodicals which are free or which must be paid for,

        - electronic resources produced by the university (teaching resources, theses, etc.) are made available on the website through the use of XML management (this is a management tool that is used to put material online).

16.2.2. *The objectives and main priorities of the document information system*

Improving the services that are available to users is the main objective behind the design of the documentation portal. The main aim is to enable all categories of users to be able to carry out an improved search when using the online resources, regardless of the medium used for both research and consultation. The aim is also to introduce a real sense of interaction for the user, where each user feels that they are part of the system. It is also hoped that the user will be provided with the most relevant response, not only in relation to their initial query, but also in terms of flexibility, efficiency, autonomy, reliability and personalization.

The document information system represents a renewed vision in the field of documentation which is creating opportunities for future research departments thanks to the progress that has been made today in terms of accessing resources, the quality of the resources that are available and the services that the research departments currently have on offer.

16.2.2.1. *Offering a one-stop shop for access to the entire set of online documents*

In responding to the specific needs of each individual, the portal (which is a tool that unifies the different ways of accessing the various types of documents produced or displayed by the university) enables us to make the wide range of documents that we possess available to everyone. The definition of the infrastructures and tools that were used during the creation of the portal responds to the issues of improving the access to information, the increased access to knowledge and the mobilization of the different types of documentation available. The meta engine also makes the access to increasingly heterogenous university documentation much easier (local catalogs, collective catalogs, digital resources that must be paid for in order to access them, free resources, bookmarks, gray literature, etc.). The portal, which was originally designed for multi-database and multi-criteria searches, has an access interface which is easier and more suitable for the users to use: from one screen and only one request the user can use the entire set of networked resources of documents.

16.2.2.2. *Possessing a development tool for resources acquired or produced by the university*

Since 2002-2003 the information revolution and the progress made in modern technology have been encouraging the research department to change its ways of managing information resources by taking into consideration the evolution of user needs. User needs are becoming increasingly linked to multimedia and also to specialized information and online periodicals (in integrated text if possible) in particular. In order to respond to user needs, a voluntary policy dealing with the development of this state of the art documentation was introduced. Online documentation is part of these development policies which focus on the collections of documentation available in each of the campuses of the University of Artois. The documentation that is available at the University of Artois is used for research that is carried out at the university. It is this provision of state of the art documentation which has been the main goal of the University of Artois' library since it was created.

One of the main objectives of the document information system consists of making the different categories of users aware of this new pool of documentation which is available to them (unfortunately this documentation that is available to them today is severely under-used). It is also the task of the document information system to make sure that this new pool of documentation is as user-friendly and

efficient as possible for the users. The document information system also enables users to have personalized access to the resources which have been acquired by the research department at any time of the day or night no matter where the user might be accessing the resources from. This is also a way to highlight the richness of the resources which the University of Artois has to offer.

Thanks to this amazing development tool for resources, and online resources in particular, the research department at the University of Artois wants to become more involved in the active production, display, development and conservation of the university's scientific and pedagogic publications:

– research resources (communications, pre-printed publications);

– teaching resources;

– Master's degree dissertations;

– resources published by the Artois University Press:

    - at the moment consideration is being given to the idea of displaying specific content online: books and articles that are out of stock, or conference proceedings which cannot be integrated within a paper publication but which can be published online;

    - the documentation will be placed in the portal (subject to authorization), it will then be referenced and indexed using the document management system. Once this process has been completed the documentation can be accessed for research using the portal;

– resources relating to culture.

The research department has a particular method of discussing and forging renewed contracts with teachers and researchers. Within the framework of teaching or research these teachers and researchers who have had their contracts renewed are asked to produce material for the portal.

16.2.2.3. *Offering personalized services to the different categories of users*

The creation of personalized services online aims at analyzing and grouping together the content that is requested by users and the content that is available to users. This aim also falls within the university's global policy on documentation. This personalization of services involves the efficient management of authorized content and access rights, and for this efficient management to take place the LDAP index is used (this index is shared and used by all departments and services of the university).

As a tool that is used for welcoming and orienting users, the portal was also designed and developed for an autonomous approach to research and to act as a support to training and self-training the user on how the document information system functions. In this way, the portal is a real media tool that helps make researching information a lot easier, therefore making the portal a real assistant to the user. The portal also makes it easier to share all of the information the document information system possesses with the users.

In order to create the portal, a profound shift in the principles and priorities forming the basis of the services we had to offer was required. This took us through three successive stages of our concept: going from a system-oriented approach that was centered around the devices which made the collections of works available to the users, to a user-oriented approach that was centered on service. We have now arrived at a third user-oriented approach that is related to the competence of the users and the type of information provided (to give meaning to the man-made tools and to convey knowledge). From here on the users construct their own information depending on their individual needs and expectations and they also determine the resources that they want to use (when carrying out their research) from all the resources that are available to them.

If the document information service provides all these resources to the users, it means that training the users on how to use these resources becomes paramount so that they can work in the best conditions possible when it comes to accessing the documents and knowledge that has been made available to them. Teachers and researchers, as well as students, need to become aware of the information available and must to be trained on how to use the resources so that they can have a general overview of all of the documents and services available to them.

16.2.2.4. *Integrating more documentation within the universities on both a local and regional level. Integrating documentation within the universities' scientific and pedagogic policies*

The implementation of the document information system was also seen as a challenge for the entire research department team. The research department questioned the role and the main priorities of such a system. It was necessary to move from a document-centered approach which deals with how well the documentation is presented, to an approach which is more user-centered by welcoming the public, analyzing the services that are available to the public, as well as analyzing the necessary technical organization of the documents to respond to the expectations of the users.

This tool was designed and developed in the framework of a collaborative and participative approach with the aim of enabling the users of the system to become active partners in the composition of the system's content. The tool also provides the

other services of the university with a real development platform that they can work with. All of these other university services also contribute to the production of documentation in the fields of:

– administration (studies, reports, surveys, etc.);

– research (publications, theses, etc.);

– teaching (programs, handbooks, distance learning, training, information and orientation and professional integration, etc.).

In this respect, the different services of the university were encouraged to work together in order to overcome the issues of the interoperability and upgradeability of mobile tools, as well as interdisciplinary and cross-disciplinary issues within the university.

There are many different publishers of documentation spread throughout the university. The coherence and secure co-ordination of the different documents involves a process of integrated construction of the tools used in the publication of documents to guarantee the user-friendliness and efficiency of the global system made available to the users.

In using web technology it has become possible to carry out more joint projects with other services of the university, for example in the development of teaching products or in the development of multimedia classes. The documentation that is used in the development of such resources has become more dynamic thanks to the interconnection of the document information system with the university library's catalog. This interconnection with the library catalog makes it possible to immediately find out what printed documents are available, and it also provides access to the electronic versions of the documents.

The software tools of the document information system are used to develop professional portals which can be integrated into the university's general information system: the Artois University Press portal, the cultural life portal, research portals, the portal of the Maison des Sciences de l'Homme (a French foundation which aims to promote the study of the humanities) for the region of Nord-Pas-de-Calais, etc.

Once integrated within the university's general information system, the document information system is the pivot of the set of portals whose interconnection aims at offering users the widest choice of resources available when it comes to carrying out research.

16.3. From the idea to reality: the spread of the document management system and the documentation portal

16.3.1. *Technical configuration of the document information system*

The document information system relies on an integrated software package which provides the system with both the functions of the documentation portal and the document management system. The interface of the document management system is found within the full text integrated library system:

– the functions of the document management system make it possible to reference the online resources, to index the online resources, and to carry out searches in full text;

– the documentation portal has a single entry point which enables access to the document information system and offers:

- federated searches on all resources (OPAC, online resources, periodicals, electronic resources, etc.),

- personalized services for each user according to their profile,

- a library intranet.

Constructed as part of a future information system for the university, the document information system uses the LDAP index to identify and authorize access to services and resources so that the system can be used no matter where the user is accessing the services and resources from. The system relies on a normal and open infrastructure which respects the laws of interoperability.

16.3.2. *The document information system as a development tool*

The document information system was developed with the aim of managing and developing the electronic resources produced within the university in order to become the sole access point to all types of documents produced by the university, irrespective of what university service produced them (teaching, research or administrative documentation). In order to develop these resources, it is necessary to make the users of the system and everyone within the university aware that they are available within the research department's collection of resources. The documentation portal thus becomes an essential chain in the research department's communication with the other services of the university, with the aim of involving the university community in the creation of a policy on documentation. The documentation portal also makes it easier for the other services within the university to see the research department as being the project manager of the portal and the

document information system, as well as seeing it as a reference point from which everyone can receive the best information possible.

The tools that make up the document information system were designed to be shared and optimized within the university. In order to provide tools for publishing within the university and to create an exhaustive database of the university's publications, it was possible to rule out two types of technical solutions.

The first was the use of the document information system's platform which evaluated the documents produced by the university within the documentation portal:

– the supply of documentation via the documentation portal published by authorized users (teachers, doctorates, etc.);

– the referencing and conservation of documents within the document management system;

– the document consultation available in the documentation portal. This could be achieved in two ways: either by carrying out a specific search for the required document or by simply flicking through all of the documents and finding the required document in that way;

– HTML versions of documents and the publication of different services on Internet sites.

The information content could also be indexed by search engines which are available on the Internet.

However, an alternative was proposed to the different services which produce the information content, i.e., using the same software as that used in the document information system to develop the different professional portals which will become part of the university's general portal.

The documentation portal will give each university service the possibility of:

– constructing their own portal with informational content and personalized documentation;

– being autonomous when it comes to supplying and publishing their content (documents, information, etc.);

– managing the distribution rights of the documents and information by relying on the verification of the university's LDAP index;

– referencing and indexing the electronic documents (produced by the service) in the same document management system as that used by the documentation portal,

which means that the documents are also accessible (subject to rights) to all searches carried out on the documentation portal.

For the processing of electronic documents and by taking the most significant example into consideration, (i.e. electronic theses) a model for managing the processing of such documents has been established. The objective is to introduce a workflow, from the creator of the document to the validation of the document, followed by the referencing and indexing of the document (in the form of metadata) in preparation for publication, and finally putting the document online.

The electronic theses are managed by the software program which links the document management system and the different functions of the portal by following the scenario listed below:

– placing the electronic theses in the portal;

– referencing and indexing the theses in the document management system module;

– making the theses accessible to research carried out from the portal, with the validation of each step of the workflow controlled by the portal;

– the theses will continue to be cataloged in the integrated library system and in the university's documentation system;

– a program which will convert the theses into XML format will be introduced in collaboration with the national workshop for the reproduction of theses.

16.3.3. *The services on offer when carrying out research from the documentation portal*

The portal, which is equipped with an index and search engine in integrated text, is able to carry out federated searches on different document sources. The portal also has connectors which allow access to external sources. The portal is an essential tool used for reducing the different types of possible searches and providing searches which better meet the needs and expectations of users.

16.3.3.1. *Document searchers*

Simple or advanced federated searches using a meta-engine: the main principle is to distinguish between:

– immediate access to:

- material resources belonging to the research department (such as the AbsysNet catalog),

- electronic resources produced by the university (theses, classes, reports, etc.),

- resources selected by the research department and which give access to integrated text (websites, electronic periodicals, CDROMs, etc.);

– the different possible ways of accessing the resources (library databases, catalogs of other libraries, etc.).

16.3.3.2. *Navigating through the resource lists*

Access to resources by theme enables the user to consult a selection of electronic resources and document folders.

In order to provide the best possible selection of resources, the research department wanted to work actively and in close collaboration with research units to ensure the relevance of the selection of the resources. In order to facilitate the production and updating of these resources, the documentation portal possesses a resource management function which makes it possible to control the content of the information coming from outside the university, the documentation portal can also manage simple workflows, etc.

Access to resources by type: pages of exhaustive lists by type (databases, bookmarks, digital stocks, theses, videos, etc.).

16.3.3.3. *Personalized services*

Using the LDAP index, which verifies which users are accessing the system, the portal offers interfaces, services and contents for each specific category of users.

The basis of this diversification and adaptation of services offered to users is the virtual office which enables each user to access:

– information on their reader account;

– saved searches and to display selected information which corresponds to their user profile;

– search baskets made up of different search sessions;

– messages linked to distance information services, such as asking the librarians a question;

– the status of interlibrary loans or express loans which were ordered online;

– the status of recommended purchases online;

– Internet favorites;

– the status of new documents which have just been placed within the portal and also accessing the different stages of the workflow process.

The document information system also allows for the personalization of this virtual office depending on the user's interests by adding inserts corresponding to the sources of information chosen by the user, or by the integration of RSS feeds.

16.4. The evolution and spread of the document information system

16.4.1. *Strengthening co-operation with other university services in order to gather and broadcast all of the digital information that has been produced*

By adopting web technology, the objective is to develop joint projects with other university departments.

The main development project corresponds to the online display of research documentation and to the online display of the individual modules of multimedia classes. The interconnection of the platforms with the library's catalog makes it possible to immediately find out what printed documents are available and also provides access to the electronic versions of the documents.

It is important to point out that a documentation methodology website has been in place at the University of Artois since 2001. This site was developed by professors of documentation within the university itself. Through its multidisciplinary content, the site is aimed at all students from all campuses of the University of Artois.

The content of the website is based on three key areas:

– methodological help when it comes to researching information, this help is provided in close contact with the university's library;

– help with the design of documents, which in turn enables students to easily find the information that they are looking for;

– the creation of specialized documentation products (resource index, commented bibliographies, internship reports, etc.).

The anticipated public who would use this website, and the targeted public, are all students who are studying for Bachelor's or Master's degrees (particularly those students who do not have the time to be present in class (i.e. students in full-time employment), or students who cannot access the university for other reasons (such as those students on work placements or students with reduced mobility) and also people who are in full-time employment as well as jobseekers (in relation to training

courses or for the preparation of an entrance exam), and teachers from other disciplines.

The content of the website addresses the principles of documentation methodology, training users on how to use the electronic collections of works (CDROM and specialized online databases), the basics in using information technology and communication (researching information online, help with what to do with all of the results of a search request, as well as help with the creation of presentations (tutorials for using PowerPoint, Impress, word processing for Word and Open Office), and help with creating documents: document folders, slideshows, web pages). The website also aims at helping students reach different levels in the computing and Internet certificate (C2i) in each of the relevant disciplines.

16.4.2. *The integration of the document information system within the global information system of the University of Artois*

The objective is to integrate the documentation portal in the digital work environment so that each user can access all of the university's documentation resources from their virtual office.

For the users to be able to access these resources, all that is required is one single authentication of their details thanks to the creation of a connector which enables the system to use the university's authentication server to verify the identity of the users and to save them from re-entering their passwords each time they want to access the site (SSO technology).

In order to make the document information system interoperable, web services and connectors must be developed. These two devices enable the document information system to have access to the functions of other applications (e-learning platforms, administrative applications, etc.) and vice versa.

The main objectives of the interoperability of the document information system can be defined as:

– the development of a web service which enables direct access to documents within the document management system which belongs to the document information system. The documents can be accessed from other co-portals (portal of classes available online, research portal);

– the development of a connector which makes it possible to widen the document search (offered on the document information system's portal) to training supports offered by the portal of classes available online.

In order to progress to using this type of system access, two development projects were undertaken for the modernization of the documentation and the implementation of the document information service:

– the training of staff and users on how to use the new interface and all the tools that make up the interface;

– the optimization of access conditions by:

- improving all of the computers and their hardware in order to introduce computers that can access the system within libraries. The aim of introducing these computers is to make the document information system commonplace and also to make it part of everyday research and study (all of the university computers should also have access to the library's system),

- the development of the Wi-Fi network so that laptop computers can also have direct access to the documentation provided by the university (there is a program known as MIPE or MicroPortable Étudiant, which is an initiative whereby the university lends laptop computers to students with the aim of facilitating student mobilization and access to the university's documentation from outside the university).

The aim of these development projects is to remove all of the obstacles which might hinder access to the online documentation, as well as making everyone aware that the online documentation is available to them so that it becomes part of an everyday research and study tool for users.

16.4.3. *Providing the content for the document information system*

In addition to the access conditions, a second criterion determines the suitability of the content provided: a documentation portal can only exist if there is material to put in it.

The document information system needs to be supplied with varied material from the different services of the university whose staff and students use the system.

Subscriptions to electronic periodicals are integrated within the portal through the use of federated research tools.

The evaluation of the documentation published by the university is carried out in close consultation with the different services of the university:

– the doctorate school for theses;

– the science council for research publications, and the service known as Information and Communication Technology for Education in relation to the evaluation of teaching material;

– the digitization guidance committee at the Maison des Sciences de l'Homme.

The objective of regional co-operation is to share the acquisition and publication of documents:

– participation in the digitization project of regional historical documents in close co-operation with the university of Lille III;

– participation in the project focusing on the display of digitized theses, with the hope of using some of the techniques used by the national workshop for the reproduction of theses in order to carry out the retrospective digitization of theses and the planned development of an electronic database store.

In addition, each system will need to be updated and upgraded in order to ensure the best conditions possible so that the regional portals can become integrated within the larger national ones (for example the Sudoc portal which was created by joining some of the local regional portals with some of the national ones).

16.5. Uses and feedback

Our documentation portal has been available to users for more than 18 months now. We do not have a breakdown of the analysis which shows the satisfaction levels of the users or which shows the reasons why the users use the system.

From looking at the statistics module within the documentation system, as well as looking at all the comments made during the introductory phase of the system, it is possible that the opinions and views of the users have not changed even 10 months after the introduction of the documentation portal.

8 out of 10 students registered at the university are in a Bachelor's degree program, the majority of these students being in the first year. First years often find themselves powerless when it comes to the amount and complexity of the documentation that is available to them, whether it is in printed or electronic format. This difficulty in finding the documentation and the inability to find what they are looking for, impacts the way in which the students use the different functions made available by the documentation portal.

The users then adapt and become accustomed to using the documentation portal as an Online Public Access Catalogue (OPAC). The documentation tool is primarily used for the OPAC (90% of all access made to the documentation portal was for the

use of the OPAC) which only counts for 10% of the total number of tools that the portal provides.

The majority of the research is carried out by using the quick search option. The users are attracted by the simplicity of the tool as it uses the same search function as Google. The training that is offered to users on how to use the system, as well as the development of the specific digital resources, has had little effect on the way the users use the portal to search for information, i.e. they continue to use the traditional and basic functions made available to them, such as the quick search option.

It must be pointed out that this partial use of the documentation portal resulting from the different course reading lists was also observed within the OPAC, i.e., different reading lists mean that different students only need to use certain functions of the portal. The partial use of the documentation portal seems to be linked to an inability to determine the different search options offered by the software, as well as the different methods for exploring the different collections of books in the library. The users who feel uncomfortable with a tool that does not help them clearly understand and master the search options that it provides, feel reassured by limiting themselves to the use of a repeated approach because it is known for its simplicity and its apparent efficiency. The other search options, either because of their complexity or because of their requirement for a precise definition of the documentation being searched, are often completely ignored, and considered as tools created by librarians for librarians.

These observations perfectly correspond to the way in which students access the documentation portal, which was observed at different university libraries. The observations also correspond to traditional research services such as the consultation of monographs, periodicals and the borrowing of books. The new services which are centered on the availability of online documents are still not widely available to the different categories of users who use the library. The number of subscriptions to electronic periodicals or specialized databases is increasing and counts for one-sixth of the research department's costs. However, having a subscription to an online periodical or specialized database does not provide any special benefits in terms of access to online documentation, which is contrary to the users' initial beliefs.

The documentation portal, as a key element behind the development strategy of resources, has still not achieved its goal in providing new documentation. Short-term, medium-term and long-term solutions have been encouraged within the research department in order to find solutions to this lack of new documentation and resulting lack in use of the digital resources. This causes a problem in terms of the documentation policy, especially when people have a look at the actual budget that is devoted to the acquisition of new resources. In other words, a large part of the

budget is devoted to the acquisition of new resources but not a lot of new resources have been purchased.

The analysis of how the students used the document information system also raised certain issues in relation to the identification and authentication of users within the documentation portal. Analysts noticed that users tended to connect to the documentation portal without identifying who they were (whether they were accessing the portal from computers in the library or from computers outside the library). Only one quarter of all of the users identified who they were (most likely by revealing their identity when consulting their reader account), the remainder accessed the portal in anonymous mode.

Refusing to provide user identification or preferring to remain anonymous greatly holds the users back from discovering the personalized services that the documentation portal has to offer them. These personalized services, which take into account the different needs of the users, are only used by a very small proportion of the portal's users. Making the users aware of what the portal has to offer, as well as training the users on how to use the portal, is essential more so now than ever before if the aim is to provide services for all of the different users of the portal as well as improving the quality of the resources and services available.

Another key function which was included in the documentation portal was the federated research option which combines all heterogenous sources. This function, which has not been used very much even though a lot of time is devoted to training users on how it works, also provides resources which correspond to the needs of students studying for their Bachelor's degree (general press information, legal references, economics databases, etc.).

These different observations clearly indicate that what the documentation portal has on offer does not really meet the expectations of the users, and that the provision of personalized services, as well as new possibilities for carrying out documentary research and the provision of online materials, requires a lot of work and effort to make sure that the users have access to what they need and what they want. For example, the training methods used have to be altered in order to train the users on how to use the system, as well as making sure that the resources provided are relevant to user expectations and needs.

At present, the university library catalog appears to be the tool most used by the users of the system to carry out their documentary research; 9 out of 10 users use the university library catalog which is more than the number who use the documentation portal.

16.6. Prospects and development

One of the first consequences of this project was the modification of the research department's place within the university, since the project made it possible for the research department to have more contact with other departments within the university. The research department finds itself within a cross-disciplinary framework both within the research department and within the university in general. This cross-disciplinary framework was founded on dialog, communication and the development of resources and services.

As a result of this there has been the necessary co-operation between the documentation, teaching and resource services within the university, which brings into question the traditional co-operation methods used between the different departments of the university, the research centers and the laboratories. Training, in relation to using the diverse and specialized resources, is now seen as an urgent requirement for the success of students, as well as for the quality of the first research works which have been carried out.

With this in mind, a major concern for the research department is determining the best possible way to respond to the needs of each user in developing, for example, the complete set of documentary reference resources for each area of study to help each student with their studies. The training programs are designed and developed in close collaboration with the training staff and teachers who recommend specific text books so that the users can develop their own work and study style which will enable them to perform to their maximum potential, and which allows them to build up their own individual stock of information and knowledge which will prepare them for entering the world of work.

This training process, which is aimed at helping the users of the system understand how the system and its components work, is also accompanied by an ongoing evaluation of the device in relation to the adaptation of the tools used, and in relation to the provision of documentation and content as far as the expectations and needs of the users are concerned.

Two examples of how the tool is evolving, and which respond primarily to research needs, are currently being developed. These are the implementation of open archives and a help tool enabling the teacher-researchers to use the archives, as well as the possibility of accessing the electronic periodicals from outside the university.

The research department wants to encourage the implementation of an open-archive tool within the university. The research department will work on this by carrying out the retrospective digitization of research work carried out by teacher-researchers and students, with a particular emphasis on the retrospective digitization

of defended theses from the different faculties within the university. These online documents, which represent the scientific publications of the university, will be placed within the document management system module of the document information system and will then be published and made available to everyone using the different research methods possible at the university. The research department will also offer more relevant sources that will be adapted to the different users to provide them with relevant resources and documentation to help them carry out the best research possible. This provision of more relevant resources will be carried out in two ways: first of all by advising the publishers of what content should be included in the material before it is placed in the documentation portal, and secondly by training the different readers on how to use this new pool of resources in order to help them with their research.

In order to respond to one of the most frequent user requests, the research department will integrate (as much as it possibly can) technology within the documentation system that enables users to access the collection's electronic periodicals from off campus. These new methods for accessing the documents off campus should increase the number of users using the online periodicals.

There still remains a lot of work to be done as far as the documentation information system is concerned, which will in turn offer the research department many opportunities to continue its relationship with the users of the system, which is the purpose of the research department. We hope that the mobilization and enthusiasm of the research department's team will ensure the development and maximum use of the resources that they have on offer.

Chapter 17

Do Libraries Have a Future in Academia?[1]

17.1. The control of knowledge

As a publisher dependent on the library market I have to be careful about what I say. In any public debate librarians are seen as more worthy than publishers. To quote the poet Thomas Campbell: "Now Barabbas was a publisher". At a literary dinner Thomas Campbell successfully proposed a toast to Napoleon at the height of the Napoleonic wars on the basis of Napoleon having just executed a publisher in Germany.

However, the library has had its critics. Foucault saw it as a place where all knowledge is accumulated (culture's external memory) playing a part in the post-enlightenment West's attempt to classify everything. Foucault's criticism was that for librarians classification became sufficient in itself, even condemning knowledge to a particular place where only experts can find it (2). Said was also critical of the role of libraries in stereotyping rather than understanding, for example, in adding authority to the West's false perception of the East (3).

Both, I am sure, would be delighted by the enabling power of Google, giving anyone in effect their own Alexandrian library. They would see the libraries' control over knowledge broken to the benefit of intellectual freedom and development.

As everything becomes digitized the library building, if not the role, may become redundant. As William Mitchell said: "It will not be possible to tell tourists where some Marx of the next millennium sat. All that is solid melts in air" (4). This certainly seems to have happened at GSK where 26 libraries were apparently effectively replaced by two "virtual libraries".

As journal publishing and its future have been actively debated recently in relation to open access and institutional repositories (and therefore ultimately the role of the library) I shall look at recent studies on journal usage. The process of change is well documented.

17.2. The changing use of journals

A survey of three academic libraries in Europe has shown that with the transition to the digital environment certain services, activities and costs will increase, whereas others will decrease. There was some variation between the three libraries but all predicted a volume increase in "full text accesses" and a decrease in paper content being borrowed and viewed. There was an indication that there will be slightly fewer staff but those remaining would be better qualified (5), a prediction that has also been made for company libraries.

In the USA there have been similar changes in service trends. Figure 17.1 shows ARL (Association of Research Libraries) data for 1991-2002. Some of the overall figures mask the true story. For example, although in-house usage dropped 35% from 1991-2002 it actually increased to a peak in 1996 and then started to fall. Likewise, reference transactions peaked in 1996 but declined by 11% from 2000-2001, and 7% from 2001-2002.

ARL data have often been used to highlight the so-called journals crisis. In Figure 17.2 you will notice some interesting recent variation in the trends of a decade. Inter-library borrowing and lending drop and serials purchased surges up. This is of course a result of the big deals. Greater accessibility has reduced the need for borrowing which is a very expensive process. The authors of the ARL report seem reluctant to accept the full impact of licensing deals with consortia and the resulting lower average cost per journal accessed (6).

The impact of the big deal is clearer from UK data. According to SCONUL (the Society of College, National and University Libraries) the mean number of journal titles received by its member libraries in 1993/94 was 3,976; the nearest equivalent figure for 2001-2002 is 6,489. In this period colleges have been admitted to SCONUL which will have brought down the mean for 2001-2002.

King, Tenopir and others have been observing the usage of electronic journals in the USA and the benefits of wider access to literature (7). In the three universities studied the average number of articles read by scientists increased from 150 in 1977 to 216 in 2000-3. Most of this reading is from library collections. Also, university scientists now read from a wider range of journals. In 1977 they read at least one article from 13 titles but now read from twice that number. King *et al.* offer various explanations including more use of bibliographic databases but conclude that the breadth of reading and corresponding increased use of library collections are also undoubtedly due to access to enlarged electronic library collections. Preliminary analysis indicated significant use of electronic titles not subscribed to previously in print.

King *et al.* suggest that another indicator of the usefulness of reading journals is that science faculty whose work has been recognized through awards or special recognition in the past two years tend to read more than others. For example, those whose work has been recognised averaged 258 readings per year while others averaged 203. And a final point: "articles read from library collections tend to be of greater usefulness and value than articles obtained from other sources".

17.3. Will the serials librarian survive?

Journal publishers by working with librarians to establish new pricing models based on online delivery of content are achieving what Andrew Odlyzko predicted in 1999 (8). Most journals are now available online, which is the necessary first step in the change in the relationship between publisher and library.

Odlyzko's analysis focused on library costs. He pointed out that internal operating costs (overheads) are two to three times acquisition budgets. The scholarly journal issue is really a library cost crisis. Odlyzko agreed that unnecessary library costs are far greater than those of publishers which create an opportunity for the latter to exploit and thereby retain their position. If there is competition for resources between libraries and publishers it will be publishers that come out ahead. Odlyzko summarizes his case:

1) There are fewer publishers, making it easier to develop large-scale electronic databases.
2) Publishers are more used to competition.
3) Publishers control copyrights and thus digitization of old material.
4) The publishers' target is more interesting: librarians have at least twice as much funding as the publishers' revenues.

Point 3) relates to Odlyzko's suggestion that archiving by libraries is an unnecessary and huge expense. Once the publishers have digitized their backlist (now virtually completed by the larger publishers) much of the cost of archiving and inter-library lending will drop out.

Odlyzko accepts that besides retaining their competence as information specialists, librarians are well positioned in two roles: negotiating electronic access licenses and enforcing access restrictions. If, as Odlyzko predicts, bundling and discriminatory pricing (8) evolve from the current "big deals", then these two roles become complex and key to the information chain.

Several years ago I wrote a short article with two colleagues, Ian Bannerman and Allen Stevens (9), comparing journal pricing models to buying wine. Document delivery was buying by the glass, a subscription was by the bottle and a site license was by the case. We likened self-publishing (now called open access) to home-made wine. We could take this analogy further and liken the library to the pub. Certainly some students find it easier to study amongst others in a library just as some prefer to drink in a bar. Obviously the library has a role as a place to study but like pubs the clientele seems to be diminishing every year. It is easier and less expensive to drink at home. Librarians know the reluctance of faculty to walk far to the library.

Perhaps libraries as buildings will become study centers with features of an Internet café. The role of the study center manager might evolve to simply creating an attractive environment for students to study in. This might even be outsourced. Two or three study centers on a campus might compete for students who would use the computers to access content acquired by the university, perhaps using specialist purchasing officers in the library or outsourcing this job to library consortium managers or agents such as Content Complete who are acting for JISC in the UK.

The journal archiving and inter-library lending role could vanish. Much of the material would be available under license with the publisher while some older articles might be accessed from services such as JSTOR or those run by national libraries. The remaining articles not covered by these arrangements might be bought "by the glass" directly from the publisher and certainly at Blackwell Publishing we have seen a dramatic upswing in electronic document delivery.

17.4. Towards a more efficient system

The librarian would still be left with a huge and complex job, for example, supporting distance learning and the e-university concept, integrating information systems and responsibilities for running the Institutional Repository, making available the work of faculty over the Internet using OAi (Open Access initiative)

protocols. The new model scholarship creating a variety of websites and other desktop digital objects on campus that fall short of "published" but are worthy of access in the future will become a major responsibility of academic libraries (10). We have great collections of authors' correspondence and papers. Who will ensure we have any record of the development of ideas by some Marx of the new millennium?

In my original article (1) I scarcely mentioned Open Access as I was asked to consider libraries. To quote Alastair Dryburgh: "I have left the library out of the list of players for the author pays model, as it is unclear what its role would be in an open access world" (11). Obviously if journal publishing were to switch entirely to Open Access then the roles of purchasing and storing hard copy would drop out.

If we take a wider definition of Open Access to include data sharing, then we can identify a new role for libraries as outlined below under section 17.8.

There has been criticism of publishers' pricing by librarians and I cannot resist the temptation to suggest "physician heal thyself". The business community is driven hard to perform better each year . KPIs (Key Performance Indicators) are part of our lives. We should like to see the library community publish measures of success/efficiency that justify their funding. Traditional measures of success, such as volumes borrowed or the number of people passing through the turnstile, are becoming increasingly irrelevant. Some KPIs could be used with publishers to set common goals and measure our joint effectiveness in reaching these. Publishers have gone through radical change in the last five years with consolidation and outsourcing giving greater efficiency but with the loss of many jobs in the UK. Is the library community prepared to go through similar changes to achieve national targets?

I hope that I have shown that through change journal publishers have made their material much more widely accessible. If King *et al.* are right we could see the UK's whole scholarly communication system perform even better as a result. We are already doing well as shown in Tables 17.1 and 17.2. If we take Robert May's analysis of citations per £m spent on research as a national KPI we can conclude that we are outperforming the USA and Japan (12) with a score of 93.2 citations per £1 m spent on R&D.

The real resistance to change perhaps rests not in publishing or libraries but in faculty. Librarians have often commented to us that they would like to move to online only but this is resisted by some of their faculty who prefer to use the printed edition. Surveys continue to show that many readers still prefer hard copy (13). A survey by the Australian Department of Education, Science and Training (14) found that 60% of researchers interviewed asserted that electronic delivery had not

changed the way they publish findings. Only 20% of all respondents felt that electronic publishing is threatening the dominance of print. Electronic publishing was often seen as second rate with concerns revolving around difficulties in verifying the authenticity and accuracy of online materials. There was a surprising lack of understanding that the online version of a journal article has gone through the same peer review and quality controls as its print counterpart. Clearly, if publishers and libraries are to realize the benefits of online delivery via a digital library, then they need to take on an advocacy role together. The traditional preference for hard copy may be weakening (15) but just as publishers cannot stop producing print editions when libraries still buy them, libraries cannot stop providing editions if users still demand them.

17.5. The challenge ahead

It is unlikely that the traditional library purchasing print-on-paper journals could survive the challenge ahead. There is a direct relationship between the number of researchers and the number of articles published. Over the last 20 years both have increased steadily at around 3% per annum, as demonstrated by Michael Mabe (16).

Research funding across the world could be increasing, however. The growing economies of China and India will support more research and thus produce more papers. In the UK the R&D spend in 2004 was about 1.9% of GDP. The aim of the present government is to get this up to 2.5% by 2014, and the European Union goal is 3% (17). All members of the EU have signed up to this goal which could in theory mean an extra 700,000 researchers. If we assume 0.75 papers per annum per active researcher (18) then the scholarly communication system will need to cope with another 525,000 papers from Europe alone.

Currently the system is handling about 1.575 million peer reviewed articles: 1.125 million from 9,000 titles in the Thomson Scientific database at, say, 125 articles per journal and 0.45 million from another 9000 journals at, say, 50 articles per journal. It would seem reasonable to assume that the annual increase in papers of 3% (16) could as much as double. We should be planning for a system to handle 2.23 million articles in 2010 and nearly 3 million in 2015.

Such projections suggest that there will be strong pressure to go to e-only for journals as the only solution. Publishers have already totally digitized the editorial office, peer review, production and delivery systems. An e-only market would reduce their costs. The decision now rests with the libraries and their users. Just as libraries have to make radical changes, they are also being faced with the new responsibility of developing and running an international system of institutional repositories perhaps linked to subject repositories. The challenge is enormous.

17.6. The versioning problem

A basic assumption in the institutional repository concept is that authors will archive with them to provide open access over the Internet. Publishers have been fairly relaxed over authors self-archiving pre-prints of their articles but these do create another version to be stored for access. The 3 million articles that might be produced in 2015 becomes 6 million items, i.e. two versions of each article. Many publishers are also allowing authors to self-archive post-prints. The definition of these varies but here we shall assume a post-print is the author's version of the accepted article. If archiving post-prints becomes standard practice, the estimated 6 million items produced annually becomes 9 million.

If libraries are to cope with ongoing growth in the production of research papers, then they must work with publishers to simplify this problem.

17.7. Developing countries

Libraries have a very important role in developing countries and are benefiting greatly from the e-only system although there are technical problems (19). The library community has worked with publishers to develop services such as HINARI, AGORA and PERI (Programme for the Enhancement of Research Information) which is run by INASP to deliver journals online at little or no charge to the library. This has transformed access to the research literature and is a fine example of what librarians can achieve when working with publishers towards a common goal.

17.8. Open computation

Clifford Lynch (20) argues that Open Access is a probable pre-requisite for the emergence of fully developed large-scale computational approaches to scholarly literature. Computational technologies require scale and comprehensiveness in the literature base that they address. Publishers are already allowing search engines to access their sites to increase demand and thus make their material more economically valuable.

Lynch predicts an explosion in services that provide access to literature residing in a great range of personal, workgroup and disciplinary settings. Such services would include specialised vocabulary databases, gazetteers, factual database, text mining, ontologics and other auxiliary tools to enhance indexing and retrieval.

The basic conceptual shift is from a system based on the assumptions that the literature and data are to be read by human beings to a system accessed by other

technical systems. The potential is massive but with its realization will involve a radical change in publishing and libraries with the latter lifted to a much higher technical level and organized to allow federated searching by new technologies and to complement the workflow of researchers.

17.9. Conclusion

Librarians will continue as an important element of academia and the scholarly communications system but, as in publishing, they should forget their past, make some tough decisions and focus on the digital future. With the global increase in investment in research, raised output will put further strain on the system. Scholarly communication must evolve quickly. Publishers can produce and deliver online and are developing new pricing models in association with the new technology. The future of libraries will depend on their embracing the e-only solution and working with publishers to achieve this solution.

Rank	Country	Citations 1993-2003
1.	USA	29,859,748
2.	England	5,582,027
3.	Germany	5,249,948
4.	Japan	4,570,289
5.	France	3,782,555
6.	Canada	3,190,200
7.	Italy	2,294,754
8.	The Netherlands	1,914,576
9.	Switzerland	1,592,228
10.	Australia	1,546,457

Table 17.1. *Top ten most cited countries (1993-2003)*
(from ISI Essential Science Indicators)

Country	Share of papers %	Share of citations %	Citations per £m spent on R&D
USA	34.6	49.0	60.0
UK	8.0	9.1	93.2
Japan	7.3	5.7	43.5
Germany	7.0	6.0	34.7
France	5.2	4.5	23.2
Canada	4.5	4.5	113.7
Italy	2.7	2.1	22.5

Table 17.2. *Publications and citations (from Robert M. May, The Scientific Wealth of Nations. Science, 275, 793-796 (1997))*

17.10. Bibliography

(1) CAMPBELL R, "Libraries: do they have a future in academia – or only a past?", *Serials*, 2004, 17, 9–14.

(2) FOUCAULT M., *The Archaeology of Knowledge*, 1972, London, Tavistock Publications.

(3) SAID E., *Culture and Imperialism*, 1993, London, Chatto & Windus.

(4) MITCHELL W. J., *City of Bits*, 1996, Cambridge (USA), MIT Press.

(5) KOLLÖFFEL J., KAANDORP A., SAVENIJE B., HÖISETH T., and RATH-BECKMANN A., "Developing a cost/benefit financial model for hybrid libraries", *Serials*, 2003, 16(1), 41–49.

(6) KYRILLIDOU M., and YOUNG M., *ARL Statistics 2001-02: Research Library Trends*, 2003, Washington DC, Association of Research Libraries.

(7) KING D.W., TENOPIR C., MONTGOMERY C.H. and AERNI S.E., "Patterns of journal use by faculty at three diverse universities", *D-Lib Magazine*, 2003, 9(10), 1–12.

(8) ODLYZKO A., "Competition and co-operation: libraries and publishers in the transition to electronic scholarly journals", *Journal of Scholarly Publishing*, 1999, 30(4), 163–185.

(9) CAMPBELL R., BANNERMAN I., and STEVENS A., "Cheaper by the case", *Interlending & Document Supply*, 1996, 24(3), 4–5.

(10) SMITH A., "New-model scholarship: how will it survive?", a report published by the Council on Library and Information Resources (CLIR), 2003.

(11) DRYBURGH A., "Alternative futures for academic and professional publishing", *Learned Publishing*, 2003, 16, 265–270.

(12) MAY R.M., "The Scientific Wealth of Nations", *Science*, 1997, 275 (5301), 793–96.

(13) ANDERSON K., "Comparing print and online readership: matching perception to reality across media", *Learned Publishing*, 2004, 17, 313–315.

(14) Australian Department of Education, Science and Training: *Changing Research Practices in the Digital Information and Communication Environment.* 2003. http://www.dest.gov.au/changing, Research Practices in the Digital Information and Communication Environment.

(15) WARE M., "E-only journals: is it time to drop print?", *Learned Publishing*, 2005, 18, 193–199.

(16) MABE M., "The growth and number of journals", *Serials*, 2003, 16, 191–197.

(17) O'NIONS K., "An OST perspective", *FST Journal*, 2004, 18, 5–6.

(18) MOED H.F., *Citation Analysis in Research Evaluation*, 2005, Dordrecht, Springer.

(19) ROSENBERG D., "The current digital status of university libraries in Africa: an INASP report", *INASP Newsletter*, 2005, 11.

(20) LYNCH C., "Open computation: beyond human-reader-centric views of scholarly literatures", in *Open Access: Key Strategic, Technical and Economic Aspects* (ed. Neil Jacobs), Chandos Publishing, Oxford, 2006.

List of Authors

Robert CAMPBELL
Blackwell Publishing
Oxford
UK

Ghislaine CHARTRON
INRP
Lyon
France

Sophie CHAUVIN
Document numérique & Usages
University of Lille 3
France

Jacques CREUSOT
INIST-CNRS
Vandoeuvre-lès-Nancy
France

Pierre CUBAUD
CEDRIC-CNAM
Paris
France

Marie DESPRÉS-LONNET
GERIICO
University of Lille 3
France

Gil-François EUVRARD
General curator
University of Paris 8
France

Olivier FRESSARD
Library curator
University of Paris 8
France

Fabien GIRARD DE BARROS
Lexbase
Levallois-Perret
France

Michèle HUDON
EBSI
University of Montréal
Canada

Paul HUNTINGTON
University College London
UK

Emmanuelle JANNÈS-OBER
Pasteur CIS Institute
Paris
France

Dominique LAHARY
Val d'Oise Library Department
Pontoise
France

Corinne LEBLOND
Library curator
University of Artois
Arras
France

Nicole LOMPRÉ
CNRS
University of Pau
France

Christian LUPOVICI
Bibliothèque nationale de France
Agence Bibliographique nationale
Paris
France

Marc MINON
LENTIC
University of Liège
Belgium

David NICHOLAS
University College London
UK

Fabrice PAPY
Document numérique & Usages
University of Paris 8
France

Michel ROLAND
University of Nice Sophia Antipolis
France

Hélène ROUSSEL
BAnQ
General Broadcasting Management
Montréal
Canada

Joachim SCHÖPFEL
INIST-CNRS
Vandoeuvre-lès-Nancy
France

This page intentionally left blank

Index