DIPLOMARBEIT

Collaboration of digital libraries

ausgeführt am

Institut für Informationsverarbeitung und -wirtschaft
Abteilung für Informationswirtschaft
der Wirtschaftsuniversität Wien

Begutachter: o. Univ. Prof. Dkfm. Dr. Wolfgang Janko
Betreuer: Dr. Michael Hahsler

eingereicht von

Maria Shustitskiy

Wien, November 2004

Table of contents

Table of contents **2**

1 Problem statement and work composition **4**

2 Theoretical foundations **5**

2.1 Introduction 5

2.2 Digital libraries – definition 6

2.3 Contents of a digital library 9

2.4 Digital library vs. conventional library 11

2.4.1 Problems of conventional libraries 12

2.4.2 Benefits of digital libraries 13

2.5 Limitations and disadvantages of digital libraries 16

2.5.1 Tasks and services of digital libraries 19

3 Collaboration **23**

3.1 General 23

3.2 Reasons for digital library collaboration 24

3.3 Models for digital library collaboration 29

3.3.1 General collaboration models 30

3.3.2 Library Consortium 37

3.3.3 Union Catalogues 38

3.3.4 The institution-driven and the product driven models 39

3.3.5 Open Access 40

4 Research issues for digital library collaboration **42**

4.1 Interoperability 42

4.1.1 Interoperability perspectives 45

4.1.2 Interoperability layers - the vertical view 46

4.1.3 Reference architecture - the horizontal view 49

4.2 Descriptions of objects and repositories 51

4.3 Users 52

4.4 Collections 54

4.5 Legal issues 56

4.5.1 Cached Copies 57

4.5.2 Access and referencing 59

4.5.3 Virtual Spaces and Non-human Agents 60

4.5.4 Privacy of User Profiles 61

4.6 Economic issues 61

4.6.1 Operating costs 63

4.6.2 Revenue model 64

4.6.3 Market positioning 65
4.6.4 Market and legal incentives....67
4.6.5 Funding sources....68

5 Technical support for digital library collaboration....70
5.1 Dublin Core Metadata 71
5.2 Open Archives Initiative Protocol for Metadata Harvesting 75
5.3 Z39.50 78
5.4 The Digital Object Identifier system 80
5.5 METS....81
5.6 OpenUrl 84
5.7 RDF 86

6 Current collaborative projects....88
6.1 The Joint NSF/JISC International Digital Libraries Initiative 88
6.2 The University of Washington Digital Libraries Initiative....91
6.3 The British Library Digital Libraries program....92
6.4 The European Library Project 93
6.5 Digital Cultural Heritage Community of East Central Illinois....95
6.6 Collaborative projects of the U.S. Department of Education 96
6.7 Chinese Memory Net 97
6.8 African Online Digital Library....98

7 "Open Access" projects and initiatives 100
7.1 Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities ...100
7.2 Budapest Open Access Initiative 101
7.3 Directory of Open Access Journals....103
7.4 The Public Library of Science 104
7.5 Create Change....105

8 Conclusion and the future directions in digital library collaboration 106

9 References....109

1 Problem statement and work composition

Digital libraries gain more and more importance in the modern world. The trend is the replacement of traditional libraries by digital libraries. In order to increase the efficiency and service quality, digital libraries have to turn to collaboration more and more often. In this work we will explore the field of collaborative work of digital libraries. We will overview the reasons of a digital library for entering collaboration and the benefits resulting from it, we will also describe several models of digital library collaboration. Then we will go over to the research in the area of digital library collaboration, considering the research topics that are most important for efficient and productive collaborative work of digital libraries in the future. As the technical support for the digital library collaboration is being continuously modified, we will give a short overview of the techniques and standards etc. being used at the present time. We will also give a short description of several collaborative projects of digital libraries running at the moment or having just been completed. We will also focus on the field of “open access” initiatives and projects, as this subject gains popularity in the modern world of digital libraries. Finally, we will describe the possible future directions of the digital library collaboration.

2 Theoretical foundations

2.1 Introduction

Information preservation and transfer have always been important issues for the mankind. The pace of these two processes has a great impact on our progress. The possibility to store substantial amounts of information and provide access of large masses to it plays a big role to the formation of our community [AnVi99].

After the invention of the book-printing and until now the main form of information storage and distribution became the libraries holding printed editions [AnVi99]. The library was historically a cornerstone of scholarly endeavor and information acquisition, and unlike previous eras, ours is an age of lifelong learning [PaWr02]. Education has nowadays become increasingly important as the core of professional and personal success. Until recently, traditional libraries could still fit the need of knowledge of the humankind, and the techniques of the preservation and use of written and printed documents were well mastered by generations of experts. But as our world changes faster and faster, it cannot be satisfied by the long time it takes for a book or a magazine to be written, typed, printed, and finally put on the library shelf. The world of information systems and the ways in which scholars, students, and the general population find and use information are in rapid transformation. Our society changes so rapidly that the information materials have to be edited and rewritten constantly, and the amounts of information in the traditional form make very difficult such operations upon them, as well as other operations such as registering, searching, distributing etc. The library had to reinvent itself in today's networked society to meet these new demands [PaWr02].

After invention of the book-printing, which was the first revolution in the world of information, the second even more significant step in this direction became the

technological development, which made it possible to preserve and distribute information in the digital form [AnVi99]. This electronic form allows storing information more reliably and more comfortable and enables many new operations on information. Information is being stored in form of digital documents, and collections of digital documents form digital libraries. Instead of a building that holds books, the library is evolving into an electronic portal, to a growing global collection of digital content [PaWr02]. The doors to this virtual library are now always open, and the library's holdings come to the user whenever needed. Digital library technologies and practices have been developed so they are within reach of every type and size of educational institution. Today's library includes sophisticated tools that make it easy to find the best information resources, delivering them to one's desktop or mobile computing device easily [PaWr02].

The essential growth of the activities in the field of digital library development was to be seen in the early 90s, when technological and informational means of secure preservation, operative processing and effective use of large volumes of information in different formats were invented. It was at this time that projects of digital libraries were being prepared in a number of countries. Some of them served as basis for the existing national and international programs [AnVi99].

2.2 Digital libraries – definition

We should define the term "digital library" at the beginning of our work. The definitions are numerous. Referring to Leiner [Lein98], the spectrum of possible meanings for the term is "ranging from a digitized collection of material that one might find in a traditional library through to the collection of all digital information along with the services that make that information useful to all possible users". Here are some definitions:

- "Digital libraries are organized collections of digital information. They combine the structuring and gathering of information, which libraries and

archives have always done, with the digital representation that computers have made possible." [Raja02].

- “A digital library is a collection of information that is stored and accessed electronically.” [RiWa98].

- “A digital library is the technologies, tools, resources, and practices associated with the management of content in an electronic information environment.” [PShJ96].

- "An informal definition of a digital library is a managed collection of information, with associated services, where the information is stored in digital formats and accessible over a network. A crucial part of this definition is that the information is managed. A stream of data sent to earth from a satellite is not a library. The same data, when organized systematically, becomes a digital library collection." [Raja02].

- “The digital library is not merely equivalent to a digitized collection with information management tools. It is also a series of activities that brings together collections, services, and people in support of the full life cycle of creation, dissemination, use, and preservation of data, information, and knowledge.” [U.A.98].

- A digital library is “a managed environment of multimedia materials in digital form, designed for the benefit of its user population, structured to facilitate access to its contents and equipped with aids to navigation of the global network” [Coll94].

- "A digital library contains digital representations of the objects found in it - most understanding of the digital library probably also assumes that it will be accessible via the Internet, though not necessarily to everyone. But the idea of digitization is perhaps the only characteristic of a digital library on which there is universal agreement." [Raja02].

- “A digital library is an electronic extension of functions users typically perform and the resources they access in a traditional library.” [PaWr02].

- "A digital library encompasses two possibilities: a) library that contains material in digitized form, and b) library that contains digital material. There is a subtle difference between the two (though not of great significance). In first case, the digital content is produced by digitizing physical counterparts (e.g. paper). In the second case, the initial content itself is created in digital form." [Raja02].

- "A digital library is the collection of services and the collection of information objects that support users in dealing with information objects and the organization and presentation of those objects available directly or indirectly via electronic/digital means." [Lein98].

The elements that have been identified as common to the many definitions are that the digital library is a collection of digital information that has to be manageable, that it includes services and activities needed for its functionality and that a universal access to digital libraries and information services is a goal. The most authors agree that a digital library must comprise three central aspects: first - an availability of digital data in a large extent, second - regional independence, third – an improved access to the data through a common interface and added value services for the user [Wies00, p.18]. A fully developed digital library environment usually involves the functions of the initial conversion of content from physical to digital form. Other important elements are the extraction or creation of metadata or indexing information describing the content to enable the function of searching, and administrative and structural metadata to help to maintain other services, such as viewing, management, and preservation. Further, there must be an appropriate multimedia repository available for the storage of digital content and metadata. Other important elements are client services for the browser, including repository querying and workflow, content delivery via file transfer or streaming media and a private or public network [Wies00].

Along with the term "digital library", there are relative terms "electronic library" and "virtual library" [Wies00]. The term "electronic library" appeared in the mid - 70s, the term "virtual library" emerged later – somewhat around 1980, and the

term “digital library” is the youngest and at the time the most popular one [Wies00]. The notion “digital library” gained its importance in the tight coherence with the World Wide Web [Wies00]. In most cases, the 3 terms are applied fully synonym to each other. As to the differences between these notions: electronic library is in some cases seen as a part of the digital library in the way that an electronic library is a collection of materials in a digital form. A virtual library can also be seen as a part of a digital library, as it is a library that is independent of any locations [EnFe00]. Sometimes an electronic library is seen as a preliminary stage of a virtual library. A virtual library is mostly comprehended as a library without walls. An electronic library can consist of both electronic and conventional part, or just be a library with computer assistance. Another term “hybrid library” is mostly used to stress the fact that the library consists of both conventional and virtual/digital parts. A restriction of the term “digital library” from the other types is very difficult [Wies00, pp. 13-18]. To avoid complicity, we shall only work with the term “digital library”, under which we shall understand a large collection of digital materials, which is regionally independent, and which offers a number of integrated digital services under a uniform user interface, and drop its synonyms.

2.3 Contents of a digital library

A digital library, just as a traditional library, has to gather content in order to offer it to its users. In case of a traditional library, the situation was clear: The library bought books, textbooks, manuals, and subscriptions for journals and newspapers, and then offered it to its visitors. The visitors of a library were usually requested to pay a certain fee for a certain period of time in order to be able to visit the library and use its materials. Information materials were stored in a library, or they could be requested from another library standing in collaboration with the first one. In this case, it took some time until the desired book arrived.

In case of a digital library, other possibilities arise. There are many sources, from which the content of a digital library might come from:

- The first possibility is similar to that of the traditional libraries. A digital library might buy a license for using a book, a journal or an article in a digital form. This is a traditional way of content gathering, and it is the most expensive one. The difference is only that in this case a digital library would not acquire a physical copy or a certain number of copies, like a traditional library. There are two possibilities: first, a library might get a digital copy of a publication. The library would then place the copy on its server and provide it for use to the registered users. The second possibility is for a digital library to obtain a right of access to the server, where the correspondent information is stored.

- The second possibility of a digital library would be to enter collaboration with other digital libraries and share the costs of information acquisition. This possibility would be less expensive than the first one. It would also be less expensive for a digital library than for a traditional library, since a traditional library has many expenses besides the costs of information acquiring, like storage and transportation costs. In case of digital libraries, there are no costs for storage or transportation of information, and so the collaborative use of materials by digital libraries is even more profitable than that of traditional libraries. The ways to acquire a work are the same as those of a single library; the difference is the sharing of acquired content.

- The third possibility is rather new, but it becomes more and more popular. This is the possibility of open access to digital information. This means that authors or researchers offer their works to the large public for free, due to various reasons. In this case, no costs arise for digital libraries whatsoever. This information access model will be described thoroughly later on.

2.4 Digital library vs. conventional library

The fundamental role and services of libraries are permanently changing as a result of advances in electronic and communication technologies. Traditionally, libraries operated as a way of sharing books [Lesk95]. Referring to Donald Simpson [Simp97] those traditional libraries focused on print-on-paper can be seen as large repositories ready to deliver information it demanded. This is called “just in case” model. The new technologies are boosting the emergence of a variety of resources, electronically locally or remotely stored and delivered anywhere in formats that transcend the limitations of print-on-paper. The library becomes a gateway for its users, and the model shifts to a “just in time” approach. This transition forces a re-examination of all facets of library methods and services. The core activity in the changing model for the scholarly process is the digitalization - coding of information for computers to process [Simp97].

It is clear that digital library technology is becoming an essential enabler of library services [PaWr02]. With the continuous technological progress and the growing popularity of digital libraries, the term itself might be heard more rarely. Just as the term “electronic banking” lost popularity and was subsumed by the more general “banking,” so will “digital libraries” be discussed less and less often as employment of the digital library technologies becomes universal [PaWr02]. The use of information technologies will be inevitable for the libraries of the future. It is certain that every library will become a digital library one day. Still, taking into consideration the great volume of material and information existing and being further created in its traditional - print - form, we must agree that print and digital material will have to co-exist for a very long time [Raja02].

2.4.1 Problems of conventional libraries

The dynamic information environment brings major opportunities as well as difficulties. As to Michael Lesk [Lesk95] the biggest problem of the traditional libraries at the moment is the cost crisis now affecting their current services. The fraction of the world's publications that conventional libraries can afford is decreasing. It is getting harder and harder for libraries to afford their previous level of purchases due to increases in journal prices, in the number of publications and due to currency shifts. Very many universities are reviewing their subscriptions and canceling journals [Lesk95].

In addition to problems buying material, libraries have to face such difficulties as increasing costs for buildings and storage. Through a constantly growing number of books there is not one library not complaining about lacking place for book storage. In order to increase this space a library would have to face further high expenses [Lesk95].

Another problem of the conventional libraries is the lack of place for readers in the libraries, especially in those libraries, where the number of students rises. This problem has always been there leading to serious inconveniences [EnFe00].

There is also a steadily growing problem of acid paper books, that is books published between 1850 and 1959, which are deteriorating and need treatment right away [Lesk95]. This treatment is very costly.

The distance which separates the library user from the library is another difficulty: the desire to visit a library declines with the remoteness. Many library users simply lack the time needed to make the way to the library [EnFe00].

Other inconvenience is the conflict between the scale of the inventory and the actual demand. A conventional library usually aims at having possibly few copies of possibly many different books. And so, if some topic becomes suddenly extremely popular, it becomes very difficult to get a desired book. The

process of ordering and registering a new book is usually very costly and therefore is another inconvenience of traditional libraries [EnFe00].

Digital libraries can be an answer to some of these problems.

2.4.2 Benefits of digital libraries

Digitalization can offer many advantages to libraries as well as to their users. The benefits mentioned by T.B. Rajashekar ([Raja02]) are the following:

- Digital libraries make it needless for the user to go somewhere. A user can get full information at home or at work, whenever there is a PC and a network collection [Raja02].
- Information can be updated continuously much more easily. It easier to keep the information current [Raja02].
- An important benefit offered by digital libraries is searching and browsing in material. One can optimize searching and simultaneously search the Internet, commercial databases, and library collections. Then one can save search results and conduct additional processing to narrow or qualify results, or click through search results to access the digitized content or locate additional items of interest [Raja02].
- Information can be shared with others more easily. By placing digital information on a server connected to the World Wide Web makes it available to everyone [Raja02].
- Duplicating of information is easy and cheap, whereas duplication of paper material would be very expensive [Raja02].
- Digital libraries compared to conventional libraries allow collaboration and exchange of ideas [Raja02].

- Arising new forms of information: information in digital form can support features and possibilities not given in print form [Raja02].

- Digital libraries are cost-saving, since expensive building, professional staff and maintenance demanded by conventional libraries not needed anymore [Raja02].

William Arms ([Arms02]) mentioned further benefits:

- Information access is not limited by geography, it does not matter, where in the world is the document located. There is no need to replicate the document because of its geographic availability [Arms01].

- The components of digital libraries are declining rapidly in price. Digital libraries are also expensive, but it is expected that digital libraries will become much less expensive than they are now, and much less expensive than the traditional libraries [Arms01].

- Possible other ways of storing information, like database or mathematics library [Arms01].

- Extended possibilities for creation of informative objects. “Even when the formats are similar, materials created explicitly for the digital world are not the same as the materials originally designed for paper or other media” [Arms01].

Albert Endres and Dieter W. Fellner [EnFe00 p. 151] found further advantages of the digital libraries:

- Information is available right away – it is a matter of minutes to get the desired title. The speed of the information access and transfer is a big benefit [EnFe00 p. 151].

- A digital document may be easily and convenient edited on a computer, if its coding and formats are known by the user, and if his computer has the required possibilities [EnFe00 p. 151].

- New comings are registered, listed in the catalogs and are available for reading mostly in a couple of hours, where as the same procedures took much more time in the conventional libraries [EnFe00 p. 151].

Finally, Michael Lesk [Lesk95] sees the following significant advantages:

- Ubiquity: a single electronic copy can be simultaneously accessed from a large number of locations by many users, assuming copyright permission is available [Lesk95].

- Loss rates by theft may be much reduced since readers get a screen display of the object, rather than carrying away the physical object [Lesk95].

- Digital libraries serve improved preservation - it is easier to copy digital information, and to do it without errors [Lesk95].

To our opinion, another important advantage is that digital libraries can contain unlimited information quantity, being not restricted by the amount of printed material that could be comprised by a conventional library.

Besides, digital libraries allow collaboration and exchange of ideas for libraries as well as users all around the globe. Being a user of a conventional library, one could only collaborate with other users of that library. Now, digital libraries can offer possibilities for international collaboration through building interest groups and offering tools, through which the users could communicate, no matter where they are. It is just as easier for digital libraries themselves to collaborate with other libraries around the globe and exchange information. A library in Austria could easily exchange digital information materials with a library in Japan. Such a scenario would be just too difficult for a conventional library.

2.5 Limitations and disadvantages of digital libraries

In the current situation digital libraries still have some limitations and difficulties. Here are some of the key issues facing digital libraries today mentioned by Michael Lesk [Lesk95]:

- An important problem of digital libraries is the technological obsolescence [Lesk95], which has two sides:
  - Hardware: the technological obsolescence of the devices for reading data is a problem, since it is a real challenge for the hardware to keep along with the changing variety of the digital objects [Lesk95].
  - Software: software obsolescence is a more serious problem than that of hardware. There are so many software formats existing and being created, that they get obsolete much faster than the hardware devices, which take a longer time to be designed and manufactured [Lesk95].
- The question of control is also one of great importance and it may lead to many difficulties. It is difficult to decide who the real provider of information is: online services, libraries, bookstores, publishers, telephone companies, telecommunication offices or university computing centers etc. In the digital environment, libraries are not the unique providers of scientific information anymore. In the nearest future, the provision of current material could move back to the publishers, who will then have the fullest possible control of information. It can also happen that, as the authors publish themselves directly to the Internet and the publishers too will be bypassed [Lesk95].

T.B. Rajashekar ([Raja02]) mentioned the following disadvantages:

- Building a digital library is an expensive process. It is not hard to collect digital resources, but it takes much more money and efforts if the original resources exist in a physical form and require transformation into digital objects. This costs decline, though, as ready library solutions are now available and one can choose the more favorable way between building a digital library with own efforts and buying an existing solution [Raja02].

- Another problem of the digital libraries is the cost of content refreshing. Digital preservation will always stay an ongoing operation, requiring considerable replicating expense. However, as the costs of technology keep declining, we can expect, according to T.B. Rajashekar [Raja02], that the cycles of information refreshing will be cheaper and cheaper. It is therefore especially important to find the funds and learn about the methods of content refreshment.

- The problem of the rights management is also very important because it is very easy to copy, replicate and distribute digital information in contrast to the same operations on a printed book. Enforcing copyright in digital environment is therefore a major issue [Raja02].

- Another key issue is that of interoperability of different digital libraries. Since the library resources are distributed all around the world, it would be wrong to expect that a single digital resource would exist. This means that we would have to think of methods to find either individual items or collections wherever they may be, and other methods to assemble virtual collections that users can browse or search through [Raja02].

A.I. Vislyi ([Visl01]) sees the following problems:

- Another limitation of the digital information is some lack of comfort which we are used to. If we buy a book at a shop, we will at first look it through and, possibly, read some abstracts at our choice. If you buy an article or a book in a digital form, you can’t look it through – all that is available are abstracts offered by the author [Visl01].

- Theft is also not abolished in the digital world, it just takes another form. Although a library may be immune from losing access to its own copy due to theft, in the digital world there is a different kind of theft in which the copyright holder loses control, rather than the library losing its copy [Visl01].

The following table offers a short overview of the benefits and limitations of digital libraries in comparison to the traditional libraries:

Benefits	Limitations
Easier and more efficient work with the information, new possibilities like searching, browsing or editing.	Lack of usual comfort of work with physical information
Ubiquity: availability of one piece of information for numerous users.	Problem of control over information
Easy, fast and cheap duplication of materials.	Problem of rights management.
Minimal loss rate of digital information by its users, e.g. due to theft.	Another form of theft: loss of control over information by the copyright holder.
Easy collaboration of the users of digital library through collaboration tools.	Problem of interoperability of different libraries and software solutions.
Cost saving: it is cheaper to maintain a digital rather than a physical library.	Costly transformation of material from a physical into digital form.
Improved data preservation.	Technological obsolescence of formats, devices and mediums.
Fast access to new comings.	
Many forms of information are	Need for refreshment of forms and

possible.	formats.
Unlimited amount of information.	
No need for going somewhere.	
Location of the piece of information. plays no role	
Easy information update.	
Fast, easy and free information transfer.	

Ultimately, we see that there are still more advantages of digital libraries than there are disadvantages. Digital libraries offer numerous new possibilities that were unrealistic for traditional libraries. The era of digital libraries has just begun, and active research in this area makes us believe that solutions to all of the now existing problems of digital libraries will be found in time.

2.5.1 Tasks and services of digital libraries

There are several ways to define the tasks of a digital library. Albert Endres and Dieter W. Fellner [EnFe00] suggest looking at them from several different points of view. From the point of view of a library user, the primary task of a digital library is to solve systematically the known problems of the conventional libraries, such as lack of place, constantly declining portion of information materials, information obsolescence or library remoteness [EnFe00]. When this task is completed, one should not forget that a solution built on a new technological basis creates new expectations and new claims, which also have to be recognized and solved, in order to make the use of the information as convenient as possible. From the point of view of the scientific research, the starting point is the technical basis available through the Internet [EnFe00]. The

task of a digital library is therefore to improve the use of the global information network with an orientation on the needs of a private user and his activity. This opinion can often be found together with the desire for a world library that can be used without the middleman function of the specialists. From the point of view of the market economy, the task of a digital library is to offer private, academic and industrial users attractive products and efficient services, which help them to get the needed and desired specialized information that is contained in the digital documents. In order to survive in the competition, a library should be oriented on the needs and wishes of the users. This service function is the dominant one [EnFe00].

Apart from these three points of view, preservation is an important matter in all areas [AnVi99]. Libraries have to place high value on preserving the irreplaceable contents of their collections. In the area of conventional libraries, fire and theft prevention systems, environmental controls or the move from acid to alkaline paper in publishing are measures against loss of irreplaceable printed items and for their long-term retention [AnVi99]. With digitized resources, former problems could be solved, but new ones are expected. Since an unlimited number of copies of an object can be created, theft or fire destruction is not so scary anymore. There are, however, a number of other problems arising, like media deterioration, evolution in type and format of media, changes in applications and operating systems and preservation of processing results [AnVi99].

The main task of digital libraries, though, is, in opinion of A.B. Antopolskiy and K.V. Vigurskiy [AnVi99], the integration of informational resources and enabling an efficient navigation among and inside of them. The purpose of a digital library is therefore to provide a central location for accessing information on a needed topic. Under integration of information resources is to be understood their aggregation with the help of a comfortable and unified user interface [AnVi99]. The aim is the use of different information under preservation of its qualities, its presentational form and the given user possibilities of its manipulation. The resource aggregation can be done physically or in the case of a digital library virtually. The main aim is to enable the user the perception of the accessible information as one information space. An efficient navigation in

digital libraries is to be understood as a possibility for the user to find the needed information in the most complete and detailed form at the lowest effort in the whole information space [AnVi99].

At the same time a digital library's task is not only to enable efficient navigation but also to provide a coherent view of a very large collection of information [LyGM95]. The many repositories of digital collections and information systems such as personal information resources, public libraries and workgroup and organizational information collections and collaboration environments should appear from the user's perspective as a united digital library system with a common gateway and a common set of support services as far as possible [LyGM95]. An even more challenging task would be to provide a coherent view of an information collection and permit information to be effectively shared across organizations at the same time retaining diversity, which naturally arises out of different purposes the information serves and organizations by whom it was provided [LyGM95].

Other requirements on a digital library are high response time, scalability and efficiency. Today, as the online digital collections of the libraries gain size up to millions of objects, digital libraries must also grow in size and offer more functions, and it is crucial for the technology to keep up with the requirements for the access and use of the digital collections [PaWr02]. According to Pascuinelli and Wright [PaWr02] "storage capacity must be scalable to adapt to rapid growth in demand, and must be adapted to the mix of media types that may be stored in a digital library, such as relatively compact text, data-intensive graphics, dynamic audio and video, which is both highly dynamic and data intensive". The expenses are also regarded and an important concern, which has to be always taken into consideration [PaWr02].

Digital libraries provide and extend traditional library services in the digital environment. [SMWB99]. A digital library can offer a number of additional value-added services, which are not offered by conventional libraries. Here are some examples, provided by Endres and Fellner [ENFE00]:

- A time-restricted access to a certain document can be provided instead of delivering a title; sometimes exactly this form of document access might be needed.

- Personalization through profile service can be offered. This service gives the user a possibility to hand over the profile of his interests in form of a list of key words. The library then uses this profile to pick out those documents out of the new comings, which may be interesting for the user. This service is also often called the Selective Dissemination of Information.

- A digital library might enable a notification service. A user can be notified per some means of communication whenever a new title – article or a book - appears in a catalog, or a new registry in a database, which matches the interests of the user.

- Creation of thesaurus and classification schemes are next important instruments for description and searching for information offered by digital libraries.

- Personal work space is a useful feature. It is adjusted for the needs of a particular user. An example could be a working space, which allows the user to use text processing, mathematical, statistical or graphical programs of a certain type.

3 Collaboration

The definition of the term collaboration we can offer is the following: “Collaboration – act of working jointly, joint operation or action” [U.A.04a]. Collaboration of digital libraries is, therefore, a joint work of digital libraries in those areas where they can cooperate.

3.1 General

The work of librarians as well as the use of library resources has been stereotyped as a solitary activity [U.A.04h]. In the literature referring to the library science and information retrieval one will hardly find any mention on the numerous aspects of collaboration. Still, it is obvious that all parties interested in extending the barriers of knowledge and information engage in significant collaboration [U.A.04h].

It is impossible for any library to be self-contained [Pede00]. No single library can ever contain all the materials and information its users can ever need. This is impossible, just regarding the tremendous and ever-growing amount of information available all around the world. An answer to this problem could be the possibility for the libraries all around the world to share the informative materials with other libraries mutually, to commonly solve problems that are beyond the ability of one institution, so that all would profit from it [AnVi99]. All libraries should have ability to search and request books and other material from other libraries when needed, or they could provide such a service to their users directly.

The tradition for cooperation and resource sharing between libraries has been there for a long time, but it existed in the form of cooperation between conventional libraries, which could borrow books from each other using

catalogues [AnVi99]. The digital llibraries will have to continue and extend this long-standing tradition for collaboration. The digitalization, the technological progress, the library automation and availability of Internet provide many new opportunities and challenges for the libraries. Collaboration becomes more effective. All those benefits a digital library may bring in contrast to a conventional library can be just as well used for the collaboration between digital libraries. Exchange of materials and services becomes fast, convenient and much more efficient [GoZa99]. A creation of a united information resource, which is like we said one of the main tasks facing digital libraries, is not the goal itself, but must serve to ease a realization of different collaborative projects. Or, the other way around, integration of informational resources into one united entity is not possible without efficient collaboration between digital libraries [GoZa99]. It was “the need to share large and complex data sets among distant researchers stimulated the origin of the World Wide Web at CERN, the great European particle physics research laboratory where the typical experiment involves dozens to hundreds of collaborators” [Simm97]. Therefore, the future of digital libraries cannot be seen without their collaboration, because this is a requirement that secures their existence and popularity.

The term “to collaborate”, comes directly from the Latin word “collaborare”, which means to labor together [Simm97]. The usual modern meaning is "to work jointly with others, especially in an intellectual endeavor" [Simm97]. The way details of how a group of libraries works together may vary depending on the aims, preferences and resources of the participating parties, but some features are common to many effective working groups [Simm97].

3.2 Reasons for digital library collaboration

Here we would like to examine the reasons for libraries to enter collaboration as well as reasons for the state or international organizations to support such collaborations.

The reasons concerning the financial benefits of the digital library collaboration were described the following way:

- **Lack of funds.** The development towards digital library collaboration is also driven by the fact that due to the growing number of published information many libraries simply lack the funds to extend their collections, not to mention development and maintaining new competences [Lesk95]. But if a library cannot keep its collection up-to-date it loses popularity and its visitors. Another big challenge for the library which needs sufficient financing is its high technological level which is needed to offer the users a collection of many different media types [Lesk95]. Whereas all this is very costly for one library, several libraries together could share their collections and software. Such a method will probably be affordable for all participants

- **Financial benefit.** Collaborative organizations of digital libraries are a way to do effective work and save money that then can be used for other library purposes [Simp97]. A large amount of money one library could spend on enlarging its collection and acquiring new software could be saved if a library decides to collaborate with other libraries, and together they could afford something they would not be able to afford before [Simp97].

- **Costly new technologies.** The trend towards creation and development of digital libraries, and the need to store, edit and give on large data volumes leads to the need of using the advanced information technologies including advances software and high-efficient computer systems. The wide usage of complicated costly technologies can be handicapped in the situation of weak funding [ZnCh02]. In such conditions an important value gets the collaboration between groups, which have already reached certain results.

The reasons important for the viability of libraries themselves and for the satisfaction of their users' needs would be:

- **Specialization.** No organization or discipline has all the needed resources to work by itself without any external help or external sources of information. In order to reach impressive results, a certain degree of specialization is needed [AnVi99]. Efficient work can only be given in case if everybody is doing his job, the job for which he or she has the needed knowledge and experience. A scientist should create new knowledge; an engineer should use his knowledge to build and improve the tools; a librarian is an information professional organize the data and teaches access to it from the point of view of a library user. The synergy required for an effective creation and use of information can only arise from joint work of people or organizations with the appropriate specialized knowledge and skills that are ready to collaborate and learn from each other [AnVi99]. Therefore, collaboration is important to allow deep information supply.

- **Professionalism.** Successful systems cannot be developed from a zero-point, but must be researched and designed together with users and domain experts [GoZa99]. There is great value when professional systems researchers and developers collaborate with librarians who bring a long-standing tradition of focusing on organization, retrieval, management and use of information.

- **Broad material spread.** The demand for collaboration and resource sharing between digital libraries has been especially growing since in addition to the growing number of published materials a number of new media's and new formats has been introduced and steadily growing too. It is would be very hard for one library to be continuously following the new pieces of information emerging and all new technologies needed to recognize and apply this material [Pede00].

- **Quality improvement.** The creation of a common resource through common work becomes a strong stimulating factor, which raises the effectiveness of overall work process, since every one single organization profits from a better joint collection [AnVi99]. We come to a situation when it becomes necessary and profitable for everyone to bring

as much to the jointly created virtual library as possible in order to improve the quality of his own individual collection [AnVi99].

- **Service quality.** Other aspect of library collaboration targets the improvement of the quality of library informational and other services, which can much more effectively be offered and used in cooperation with other libraries [Zabe01].

- **Education quality.** Raising the quality of education of all levels with the help of the modern information technologies could be another positive effect enabled by digital library collaboration. This, turn, leads to a higher level of education of a country, to a better educated population, to a raised demand for information among different groups of population and to a higher popularity of digital libraries [AnVi99].

Reasons essential from the point of view of the functionality and organization of the digital libraries are:

- **Need for coordination.** Collaboration is also important from the point of view of coordination. Due to unitary coordination every library knows what materials and titles other libraries hold or are to acquire, so that other libraries do not need to buy the title themselves, but can just obtain the title from the library holding it. It is especially as material collections get more homogenous. Coordination makes the whole system transparent and different libraries can rely on each other [Simp97].

- **Data preservation.** Digital library collaboration could be important for materials preservation - saving of the most valuable archive and museum collections by mutually creating and storing their digital copies [AnVi99].

And there are reasons that deal with the collaboration benefits on the level of a population or a country, or even the world as a whole:

- **Satisfaction of population needs.** To the digital library collaboration is closely connected the benefit of more full satisfaction of social, educational, scientific, cultural and other needs of users and potential users of all kinds through creation and usage of integrated information resources [Zabe01].

- **The international scale of demand service and information.** Because the demand for information and services can be expected to grow really large it will not be possible for a single corporation, institution or organization to provide all that the users would like to see in a digital library. [BBCC03]. This problem will grow if we consider the tremendous number of users around the world demanding content and services from anywhere. Besides, a large portion of the information and services that will be made available through digital libraries is expected to exist as a privately held intellectual property. Thus, a large number of groups will have to provide digital library content and services on an international scale and they will have to collaborate with each other to make their services more useful and attractive [BBCC03].

- **International prestige.** An important reason for the government of a country to support digital library collaboration is that through enabled access of inland and foreign users to the informational resources of the land in the areas of science, culture, education and health, the popularity, reputation and attractiveness of the country among other countries can be raised [AnVi99].

- **Investment attraction.** Spreading of information about the scientific and technological achievements of the national scientists and about the scientific-technological potential of the national companies serves the investment attraction into different projects and areas of research and development [AnVi99].

- **Language issues.** Another profit from digital library collaboration is the expansion of the use of the national language abroad and so satisfaction of the informational needs of the given language speaking users inside the country and abroad [AnVi99].

- **Role of the country in the world.** Digital library collaboration through wide information spread leads to the enlargement of the role of a country in the world informational market [AnVi99].

3.3 Models for digital library collaboration

Since librarians have traditionally turned to inter-library collaboration to solve problems that ware beyond the ability of a single organization, a number of opportunities for libraries to collaborate have evolved. The negative effect of the increasing number of collaborative organizations is that resources are being spread more and more thinly whereas coordination and management structures are needlessly replicated again and again [Simp97]. An exception to the practice of replicating cooperative organizations is the recent emergence of more local consortia using new technology for concentrated resource sharing. These are often more effective due to keeping much more in touch with local priorities and needs than national entities. So, significant role of a national collaborative organization lies in coordinating resource sharing in focused programs across many local consortia in addition to providing services that are more effective on a larger scale [Simp97].

Library collaboration is multilateral. There are many ways in which libraries can collaborate, and so many different models have been designed and are being used for the establishment and carrying out of inter-library collaboration. There often exist a number of rules and regulations in this field set up by the ministry responsible for the libraries [Pede00]. Different models can be used for different collaborative ways and for different kinds of libraries, for example it can be seen that one model of collaboration is implemented for public libraries and some another model for the university libraries etc. Different library types might also refer to different ministries [Pede00].

The digital library collaboration models are also being differentiated according to different funding schemes. The funding available for the libraries is often made

conditional to the use of the official collaboration model for this type of library, which means that a library has to respect the rules and the lending ways set in the model, otherwise it will not get any funding [Pede00]. Such a policy has a positive and a negative side - it might secure a rational and equal distribution of all the resources available, but the model can also hold back the development of those libraries who wish to expand their collaboration beyond the boundaries of the prescribed model and develop new services and facilities in collaboration with other libraries [Pede00]. In order to eliminate this negative effect certain efforts are undertaken to stimulate and support the collaboration between different types of libraries, at the same time offering the libraries more freedom to establish and extend their collaboration with other libraries of their own choice. This trend is supported by more liberal regulations on this subject in many countries as well as by overall technical progress that offers new possibilities for a simpler communication between digital libraries. Another possibility to stimulate the development towards freer inter-library collaboration would be to base the funding models on performance based indicators rather than fixed in advance determined budgets [Pede00].

There are several models or funding and cost sharing schemes existing.

3.3.1 General collaboration models

Collaboration is an issue that concerns a great number of institutions and organizations. There are some aspects of collaboration that are common to all, no matter who the establisher is. Timothy Butler and David Coleman [BuCo03] examined five general models for collaboration that vary according to their level of interactivity, based on the experience of working with many different organizations. According to them, libraries are the type of organization that applies collaborative functionality the most.

Butler and Coleman have been able to classify five primary collaboration models. The majority of collaborative environments fit into one or more of them. These general collaboration models are:

- Library
- Solicitation
- Team
- Community
- Process Support

The goal of Butler and Coleman was to help an organization determine, which model(s) of collaboration are most suiting for it. Figure 1 “Models of collaboration” that you will find below shows the five models and their relation to the number of users and the level of their interactivity.

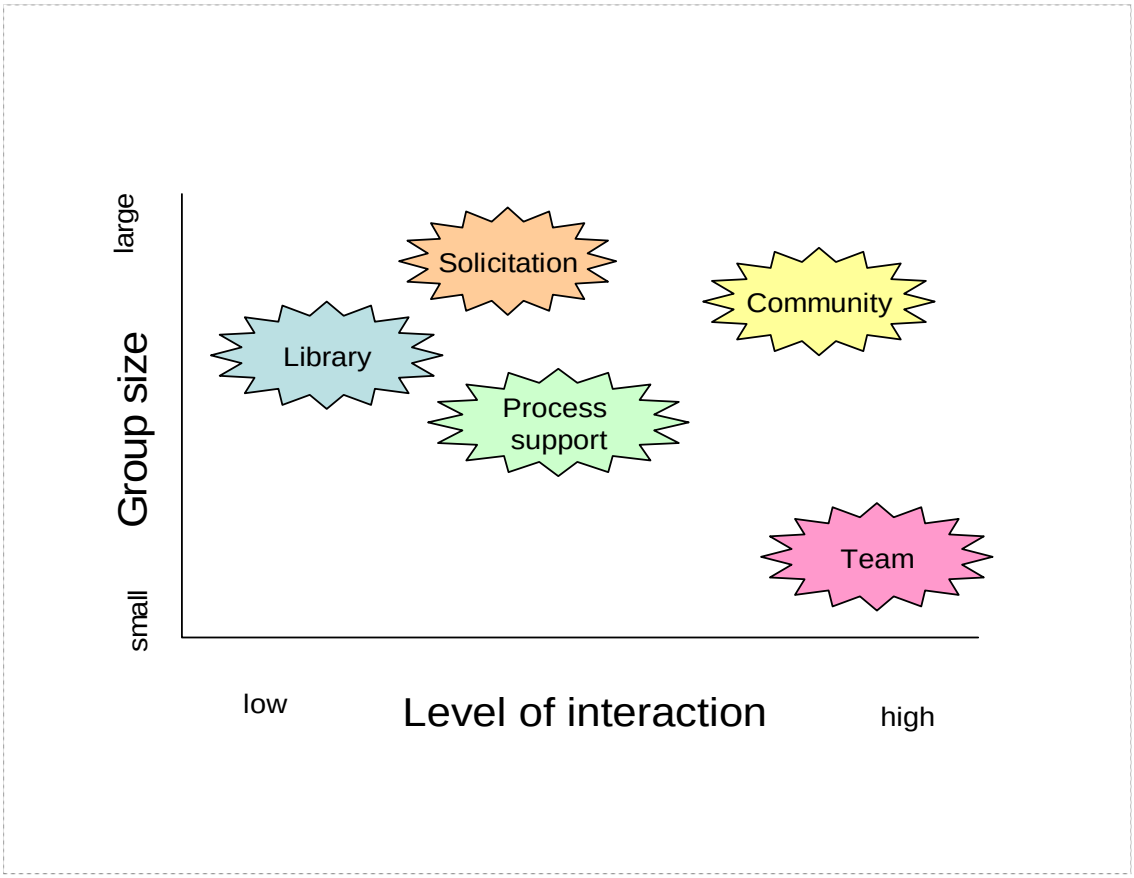

Figure 1: “Models of collaboration”. Adapted from [BuCo03]

Library model

This is the simplest and perhaps most common model [BuCo03]. The model offers shared data that can be accessed by a large number of users. The model

is not as suitable for active interaction as other models [BuCo03]. The Library model provides reciprocal access to common usually long-term use content. Here there are many more users of content than there are providers [BuCo03]. The content usually managed by a small set of people, and there is typically little or no feedback or interaction with the content creators [BuCo03]. The content is often used in asynchronous collaboration although it is possible to use it real time collaboration as well [BuCo03]. There is a complicated indexing and retrieval mechanism based on keywords, context or metadata is case of a digital content [BuCo03].

A possible use of a Library model in digital library collaboration is to offer the content of all participants for the common use. Such a Library would have a common database and provide lists and search functions to the users to quickly find the information they need. Data base would exist over a long period of time, constantly maintained and revised by a small staff group. The users would then be digital libraries, not restricted in their number.

Solicitation model

This is the next model presented by Butler and Coleman. It involves requests from a small number of requestors and numerous responses from respondents [BuCo03]. The model is characterized by a usually larger number of respondents than requestors [BuCo03]. The responses would often not know about other respondents [BuCo03]. Respondents can pose questions to requestors, and requests and responses are often moderated [BuCo03]. Collaborative interaction is almost always asynchronous and goes through e-mail or web site [BuCo03]. There is usually automatic notification to participants of new requests and responses [BuCo03].

This model is less suitable for the digital library collaboration. And still, we could imagine it working: there can be library information support group established, and the participants of the collaboration would be able to pose requests or questions. According to Butler and Coleman, standards associations use this

model to solicit review comments on proposed standards, and this would also be a possibility of its appliance for digital library collaboration.

Team model

The third model is the Team model. The model is usually used to enable the team activities [BuCo03]. It is characterized by the fact that its members share a few common objectives and have a shared stake in their success [BuCo03]. Participants are often bound by the parameters of a project and are interdependent [BuCo03]. Membership is usually restricted by a small number, like two until twenty members [BuCo03]. Membership is strictly controlled and access and security are fixed and often based on roles, groups, or projects [BuCo03]. There is a very high level of interactivity and most members both read and write content [BuCo03]. Co-editing, project control meetings and executive overviews are also common [BuCo03].

The electronic realization of the model must provide enough possibilities for members to share ideas, ask questions, and place important team documents [BuCo03]. Team members must have the facilities for discussions in order to work through problems and issues. Teams work in both real-time and asynchronously, and so the model tool must support both types of interactions [BuCo03]. “In the first two models (Library and Solicitation) security is focused on the content and access to it” [BuCo03]. In team modal the approach goes further, the interactions are considered more significant than the content and so security is being focused mostly on the interactions [BuCo03].

In the field of digital library collaboration, this model will not find frequent appliance, but it still can be used occasionally. For example, we could imagine that the digital libraries start working on some common research project and would pick out several staff members for it. This would surely lead to an active, rather than passive collaboration between digital libraries, and the need for active interaction would rise. In this case, digital libraries would find themselves in just the same position as members of other research teams using the Team model.

Community model

This is a less common, but very powerful model [BuCo03]. It is used to facilitate the activities of a community [BuCo03]. It would be defined by common interest, affinity, or goals of its members [BuCo03]. Members of the community are often self-grouping, and wish to share information [BuCo03]. Members seek to extend their knowledge and understanding of the practice or area of interest [BuCo03]. Membership is not or almost not controlled and membership must be relatively large to be self-sufficient, because new content will always be needed [BuCo03]. Communities can be moderated, facilitated or edited, all members are welcome to both read and write content, but most members are only interested in reading [BuCo03]. The number of active contributors is usually around 10% of the community population [BuCo03]. Most of the interactions are asynchronous, but over latest time “chat” communities with active interaction have arisen [BuCo03]. There are usually rules of engagement, or appropriate behaviors for the community and they are often well defined [BuCo03].

A Community model is usually organized around people who share a common discipline [BuCo03], such as e.g. Dublin Core Metadata developments. Many of the members can take part in other collaborations using other models, but, in the Community, they are brought together by their shared interest [BuCo03].

Earlier, online Communities usually offered e-mail distribution lists [BuCo03]. Recent implementations enable the advanced capabilities of threaded discussion [BuCo03]. Besides, other features can be used such as search, unseen maps, moderation and notification, particularly by the large size communities [BuCo03]. The search functionality enables retrieval of information that may be old and would be lost otherwise [BuCo03]. Unseen maps allow show members what new information has been added since their last visit [BuCo03]. If there are moderators, they can monitor postings for quality, remove unnecessary entries and facilitate responses from recognized experts [BuCo03]. E-mail notification has been shown to be a powerful tool in attracting members to the Community [BuCo03].

For the collaboration of digital libraries this model can be used as well. For example, the libraries might agree on certain technologies, protocols and standards that they would use in building and maintaining, describing etc. their collections. Then they might be always interested to find out about the development of the technical support they have chosen, about new possibilities and functions. Or they might have questions about better implementation and appliance of the standards and products they use. Then, they could just as well become members of such communities, and share their interest with others.

Process Support

This last model described by Butler and Coleman serves the use of collaboration technologies in a process or workflow [BuCo03]. Some of its characteristics are frequently performed complex or exceptional processes and self-service status monitoring [BuCo03]. The model is frequently used in conjunction with other models. The critical organizational collaborative processes that the model could be used for include new product development, sales/marketing, customer service and support, training, etc. [BuCo03]. The model offers the ability to generate customized forms that support these processes [BuCo03]. This model often offers an overview of process progress to the process manager; and there is often an underlying workflow system that supports automated transitions based on activity or time [BuCo03].

This model suits collaboration because there is usually a strong need to interact around a given process to solve problems and complications [BuCo03]. In this context, Butler and Coleman mention the Pareto's law (the 20/80 rule): 20% of resources are used to support 80% of the standard interactions, while the other 80% of the resources go to support the 20% of the non-standard, or exceptional interactions. The latter are much more complex processes, requiring clarification, rules arbitration; exception handling, etc. The model suits all this perfectly [BuCo03].

It is difficult to find an example of appliance of this model for the collaboration of digital libraries. A possible scenario could the following: one of the information materials the collaborative digital libraries offer is a protected document, offered for sale. One of the users decides to buy this product and requires it at a

website of one of the digital libraries, at the same time making a payment through the facility offered at that site. The order follows automatically to the database of another digital library, standing in collaboration with the first one. In a few days the customer calls and complains that he received the wrong product. Here, an exceptional situation arises, where the customer representative must take many steps in order to find out how this could happen, like tracking the order, calling the other library, talking to the technical support etc. This is an example of a situation, when the process support model could be useful according to Butler and Coleman.

These are the five collaboration models examined by Butler and Coleman [BuCo03]. They define some of the basic building blocks of collaboration. Many of these models can be aggregated, or used simultaneously, but separately or combined into hybrid models [BuCo03]. An example offered by Butler and Coleman is the combination of the Library and Solicitation models. This could help create a “best practices” library by soliciting feedback on each practice [BuCo03].

In order to build a framework for analyzing and designing collaboration it is necessary to develop an understanding of each model separately and then analyze their possible combinations [BuCo03]. In order to find the best solution, it is important to study the situation and make a list of the requirements. Then one could produce a plan of the available models, pure or combined [BuCo03]. Then, “by iterating through the models, additional opportunities for added value within the system can be found” [BuCo03]. Many components can be put together into a complete collaborative system [BuCo03]. A thorough understanding of the requirements is the inevitable issue if a perfect solution is to be found [BuCo03].

3.3.2 Library Consortium

The two library collaboration and funding models described by Gordon Pederson [Pede00] are “library consortium” and “union catalogue”. They regard not only general collaboration principles, but offer possible solutions for collaborative efforts in a more specific way.

The model “library consortium” is usually applied when a group of libraries often exchange of inter-library loans and/or when they use or plan to use the same software system, explains Pederson [Pede00]. A library consortium is usually set up between coequal libraries and it is often steered by a managing committee with representatives from all participating libraries. The libraries will usually share the consortia cost correspondent to a distribution key they set [Pede00]. The advantages of this model are that the libraries can use effective inter-library loan and software systems and share their costs. Shared software makes it also simpler to search for materials and order them from the participating libraries. It makes the whole collaboration with regard to the acquisition and cataloguing processes easier. Collaboration between digital libraries does not have to end with the system of inter-library loan of books. The collaboration within a consortium could go further to the system of electronic document delivery of journal articles and shared competence development [Pede00].

In respect to the collaborative digital libraries, the model offers the possibility to buy licensed digital content and then share it. This is a possibility to save expenses and to use the resources in a more efficient way.

The prerequisite for this model is fast and reliable communication inter-linking between the libraries [Pede00]. According to Pederson, the server, with the catalogues and other information, is often placed at one of the participating libraries and other libraries access this server via the net. Many modern IT systems are based on a client/server model with graphical user interfaces and demand a high capacity link between the server and the computers in the participating libraries [Pede00]. The Internet offers a good solution for library collaboration by making it possible for the libraries to be able to establish

individual Internet connections, according to their specific needs, and to be able to upgrade these connections individually [Pede00].

3.3.3 Union Catalogues

According to Pederson, many libraries have set up another form of collaboration – collaboration on the basis of one or more common catalogues called the union catalogues. Such union catalogues are often established by the initiative of the ministry or other authorities responsible for the libraries [Pede00]. As different types of libraries e.g. university libraries and public libraries often refer to different ministries it is quite normal to have different union catalogues for those different library types. The union catalogue will normally include bibliographic information and holdings of the participating libraries that are available for use and sharing [Pede00].

A union catalogue is normally established and maintained by a separate organization, which is either owned commonly by the participating libraries or is it owned by the authorities responsible for the libraries, says Pederson. It is normally implemented via a central database or - and this is used more and more - via a cluster of separate Z39.50 enabled catalogue servers, where all participating libraries have access to this database or cluster of servers [Pede00]. The participating libraries can then search and retrieve information for their local catalogue and they can search and locate material in the other libraries [Pede00]. It would be also possible to use the Open Archives Initiative Protocol for Metadata Harvesting, which will the described later on.

In regard to the digital library collaboration, the model seems to be possible as well, but less effective that the “library consortium” model, because here the libraries do not coordinate their efforts with each other and do not have to use the same software, a fact which could lead to the problem of interoperability. The libraries can only search and request materials owned by another library. The inter-library loan system is here not as suiting for active collaboration as the “library consortium” model.

3.3.4 The institution-driven and the product driven models

Rodney Perry [Perr02] offers another classification of library collaborative models: he splits the collaborative models into traditional or institution-driven and, on the other hand, innovative, or product-driven collaborative models.

The institution-driven model of collaboration is an effort to combine dispersed regional assets into a uniform database. Work activities are planned and conducted separately, and the separate work is then integrated into a uniform database [Perr02]. This model brings several difficulties with it: such as “the varying schedules and expertise of the various partners and the lack of direct responsibility for creating and developing educational products” [Perr02]. This approach requires complex project management activities to integrate various disciplines, grant requirements, schedule and production cycles, staffing skills, and workflow coordination [Perr02]. This is a de-centralized model, and so, the main work has to be done in the area of project coordination and communication between partners, instead of the project area itself.

The second approach, the product-driven model, is to be used for the accomplishment of the evolution of digital collections into digital libraries [Perr02]. Here, a management authority is to be established to control activities among institutions. The authority is then responsible for the product creation [Perr02]. The user is here seen as an integral part of the collaborative production units. A new hybrid knowledge base is usually built within the collaborative work unit [Perr02].

3.3.5 Open Access

The new and modern model of information exchange, which gains more and more popularity, is the model of “open access”. Open access is being defined by the Budapest Open Access Initiative as the “free availability [of the literature] on the public internet, permitting any users to read, download, copy, distribute, print, search, or link to the full texts of these articles, crawl them for indexing, pass them as data to software, or use them for any other lawful purpose, without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. The only constraint on reproduction and distribution, and the only role for copyright in this domain, should be to give authors control over the integrity of their work and the right to be properly acknowledged and cited." [CCEF02]. “Open Access removes price barriers (subscription, licensing fees, pay-per-view fees) and permission barriers (most copyright and licensing restrictions)” [CCEF02].

Open Access is possible due to the wide usage of Internet. Internet is the medium enabling the free and unrestricted access to the peer-reviewed literature world-wide [CCEF02]. The goal of those supporting the Open Access model is to support research, intensify learning and provide equal access to information for everyone [CCEF02].

In order to implement the Open Access economically, an alternative business concept is required. One of the possibilities is to “treat the costs of publication as the final integral step of the funding of a research project” [CCEF02]. In comparison to the expenses of the conventional forms of information publishing, the costs of enabling open access are far less [CCEF02]. There are many ways to cover these remaining expenses such as through “foundations and governments that fund research, the universities and laboratories that employ researchers, endowments set up by discipline or institution, friends of the cause of open access, profits from the sale of add-ons to the basic texts…” [CCEF02].

There are many initiatives and projects arising in the latest time, such as the Budapest Open Access Initiative, the Berlin Declaration on Open Access to Knowledge, Create Change etc., which collaborate with each other and actively

support the Open Access model of information dissemination. There are also plenty of Open Access journals, for example issued by the Public Library of Science or included in the Directory of Open Access Journals. The tendency shows that even more Open Access initiatives will arise in the future and that the number of open access journals publishing peer-reviewed literature will grow with every year.

In respect to the digital library collaboration, Open Access can be used as a model, where digital libraries produce works and article either alone, or in collaboration with other libraries, and then offer it to the other libraries or readers for free use.

4 Research issues for digital library collaboration

Most of the activities that are undertaken on the subject of digital library collaboration deal with the area of research and development. There are many topics in this field that need to be explored in order to find common efficient and useful solutions for inter-library collaboration. After this research phase is mostly completed, the time of implementation and wide deployment of its results will come [Scha95]. There are numerous topics for research on digital library collaboration and it is important to identify those research areas that should be effectively addressed through international collaboration [NWDH96]

Here, we will regard some of the research issues, such as the syntactic level of digital library collaboration, or interoperability ([Scha95]), the semantic or descriptive level ([Scha95]), the questions of user interfaces and of the collections management. Not less important is the legal aspect as well as the economic and social issues. Technical support is an inevitable research topic, since the technology partly created many problems of the research topics, and any solution to any problem will have to result in new technological developments [NWDH96]. Research in all of these areas will be needed if digital libraries are to be successful.

4.1 Interoperability

Interoperability of digital libraries is their capability to exchange and share documents, queries and services among each other or within various components of a single digital library [BBCC03]. It is an important issue for digital libraries as it is the ability to generate a single virtual view on many

different libraries or library components without them have to give up autonomy and without their physical merger [BBCC03]. The objective of interoperability is to share digital library sources and services.

From the point of view of computerization, many of the most difficult problems in digital libraries are aspects of the interoperability issue: the ability to make a multiple computing systems to work together [Arms99]. It covers a variety of topics, from syntactic interoperability: interoperability that provides a superficial homogeneity for navigation and access, but relies almost entirely on human intelligence for coherence, to semantic interoperability, where separate computer systems come to a common understanding of the information at the computing level [Arms99].

Interoperability is a central issue in the development and maintenance of digital libraries as it deals with the global architecture necessary to deploy digital libraries widely. The interoperability level deals in the first place with the techniques for passing digital objects and operations around the network between libraries, collections and users. These issues focus mostly on naming of digital objects, protocols for object transmission, types of digital objects and metadata at the syntax-level - registering and reconciling the object schema [Scha95].

The interoperability of digital libraries is an important topic because collaboration demands, beside the will to collaborate, the support of sufficient IT facilities. The establishment of the IT support is the most difficult and costly process among others and is therefore usually the last stage to be automated. The libraries have to agree about the ways to collaborate, the standards to be used and so on. In order to communicate the participating libraries must either use the same software or establish communication facilities between different software systems. The main difficulty is the involvement of other libraries and the need of integration with other library systems [Pede00]. The integration of IT systems has to be planned carefully and the agreements with other libraries have to be met as soon as possible so that the software support can be based on those agreements. The software systems for library collaboration should be planned and implemented step by step. The first step according to Gordon Pederson [Pede00] is to share catalogue and holdings availability information.

This is normally the main purpose of a union catalogue mentioned earlier. The following stage of the IT support implementation is the implementation of the inter-library reservation and lending system - the Inter Library Lending process (ILL), which is either managed through integrated in the union catalogue central managed functions, or is established directly between the individual libraries [Pede00]. The functions for search and ILL should be integrated. The implementation of the supporting IT facilities can be done in different ways and through different technologies and standards, which have to be resolved in the first place. The use, implementation and testing of different standards is an important complicated and time consuming topic for discussions, which has to be approached seriously [Pede00].

The issue of interoperability is very important, since there are so many independently managed digital libraries being created all over the world. Every library has different management policies. Computing systems of digital libraries are also different: some are modern and high-tech, others are mature and are not popular anymore [Arms99]. Due to this variety of computing systems around the world, collaboration of digital libraries can only be achieved by ensuring functional interoperability.

Interoperability and standardization are interrelated matters [Arms99]. Development of common standards is essential for establishment of interoperable systems in digital libraries. However, the formal process of creating international standards is often the opposite of what is necessary for interoperability in digital libraries [Arms99]. The official process of standardization is too slow for the rapid developments of digital libraries. Many of the standards encouraged by the officials are excessively complicated [Arms99]. A lot of international standards have never been tested in real life. In practice, if a standard is applied by a large group of users, it automatically becomes popular [Arms99]. “Sometimes a de facto standard emerges because a prominent group of researchers uses it” [Arms99]. Some standards become popular because the leaders of some large community choose to follow certain conventions. Sometimes, generally accepted standards are created from a formal standards process. Other de facto standards are proprietary products from prominent corporations like Adobe's Portable Document Format (PDF) [Arms99].

One of the main topics of the interoperability issue is its level concerning data interchange [POLO00, p.163]. It is seldom that the same software is used to create and then read the created data. Since we usually read data with the help of some other program rather than the one used for its creation, the rival software might generate some problems. Therefore, the data is often exported into a data interchange format and then transmitted into the rival solution. Some of the data or data quality will be lost during this process, depending on the level of interoperability between these applications. The danger of data loss grows with the increasing complicity of the digital documents. Therefore, complex digital resources need high level interoperability systems [POLO00, p.163].

As interoperability mechanisms will form an infrastructure, rather than a variety of proprietary components, it is particularly appropriate that the research and development of this publicly usable infrastructure be undertaken with public rather than private resources [BBCC03].

The topic of interoperability can be seen from different points of view. There are several ways to illustrate the interoperability issues.

4.1.1 Interoperability perspectives

One way of seeing the interoperability components is to spread them is a sort of a spectrum, the way Margarete Polok [Polo00] did. At one end of the spectrum is the use of common tools and interfaces that provide a superficial uniformity for navigation and access but rely almost entirely on human intelligence to provide any coherence of content. At the opposite end of the spectrum is deep semantic interoperability, which deals with the ability of a user to access, consistently and coherently, similar (though autonomously defined and managed) classes of digital objects and services, distributed across heterogeneous repositories. Deep semantic interoperability is a very important

research problem. An intermediate position between these two extremes advocates primarily syntactic interoperability - the interchange of metadata and the use of digital object transmission protocols and formats based on this metadata rather than simply common navigation, query, and viewing interfaces, as a means of providing limited coherence of content, supplemented by human interpretation.

4.1.2 Interoperability layers - the vertical view

Interoperability comes in many forms. The “EU-NSF Digital Library Working Group on Interoperability between Digital Libraries” [BBCC03] sees interoperability occurring at different levels of abstraction, which can be referred to as interoperability layers. From this perspective, referred to as vertical view [BBCC03], interoperability is the definition of languages, with requisite syntax, semantics, and pragmatics, to connect digital libraries. These languages will be defined at various levels of abstraction, and will be specialized for particular uses. It is expected that not a single language for interoperability will exist, but a variety of them. In this regard there are two tasks to be solved. An effective collaborative functionality has to be supported through common languages and tools which, on the other hand, have to be general and extensible enough to permit or facilitate system evolution and development.

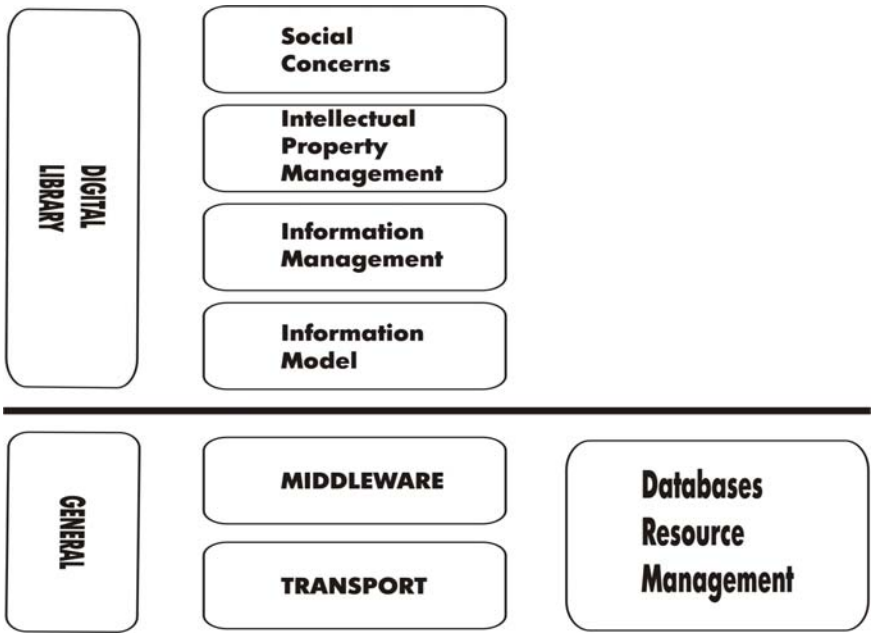

Figure 2 “Interoperability framework from the vertical view”. Adapted from [BBCC03].

In the figure 2 we see the vertical view of an interoperability framework. It builds from the transport of bits along a network, to the commonly called 'middleware', where data structures, distributed objects and various protocols (e.g., http, CORBA, Z39.50) are used to format and deliver bits to the proper place [BBCC03]. The concern at this point is reliable delivery of packets, and is not restricted to digital libraries. The transport level defines common mechanisms for accessing collections across the Internet.

The next level is the middleware level, where the framework is concerned with transforming packets into the basic elements of digital libraries: documents of different formats, queries, and other related items, and therefore differs from other network applications [BBCC03]. This level includes formatting, searching, and displaying. The goal of the middleware level is to provide a higher level of abstraction for interoperability. The different middleware standards specify what an object is and how programs interact with those objects. Distributed object models are a common way of describing distributed systems. These models define interfaces to objects in the system and provide mechanisms for communicating with these objects across machine and network boundaries

[BBCC03]. There are a number of competing standards for distributed object systems and proposals for the next generation of World Wide Web protocols - recommend the use of objects and operations on objects to define access to WWW content [BBCC03].

These first two levels refer to the general database resource management. They are not digital library specific, but are given in any collaborative project or collaborative procedures, which require interoperability.

The following levels are mostly typical for digital libraries.

At the next level – the information model level, there are definitions of document and query types - what particular objects exist in digital library systems and what are the relationships between these objects [BBCC03]. A digital library will use query objects, collection objects, and document objects. An example of such an object model is the Document Object Model proposed by the World Wide Web Consortium [BBCC03]. This object model defines the structure and operations for document objects where the document is SGML encoded. This level includes also concepts of document descriptions, and mechanisms for delivery of digital library content and services [BBCC03].

The information-management level is building upon the information-model level. At this level, the concern is not on composing or transporting information like documents, but on its proper management [BBCC03]. Here documents and services are being treated as atomic. Examples of information management include choosing a variety of services to properly process a document, e.g., translate it from English to German, dither the color images to black and white, format onto DVD, and filter email messages. By the nature of these tasks, it is expected that Metadata will be important [BBCC03].

The next level is the intellectual-property-rights management level. At this level, the prior concerns are those of property rights and the actual documents and services can be safely abstracted [BBCC03]. Examples of the activities at this level include negotiation, contracts, and various economic issues. On the level of intellectual property, services tend to be of relatively long duration compared to services on the level of information management. For example, a service

might be to "negotiate access to a collection", compared to "answer a query", respectively [BBCC03]. Similar technical infrastructure, however, might support advertising and finding services on each of these levels.

At last, there is the social-information-management level. This level, like the intellectual-property-rights level, is concerned with various legal, social, and cultural as well as with broader information and service-use issues. Language to support interoperability on the level of social concerns will focus on representing the needs of users [BBCC03]. An example of interoperability at this level is the PICS standard, which enables web-sites managers to rate their sites on various content categories, such as violence, to allow possible users to be aware of what the site contains. Metadata on other levels of the interoperability hierarchy describes digital library content and services primarily from the perspective of system functionality rather than user needs, like here [BBCC03].

This vertical perspective gives us a notion of the interoperability components, but it does not capture the dynamic aspects of interoperability in digital libraries, such as query processing and management topics. An answer to this problem is reference architecture, which represents the second - horizontal perspective to interoperability issues [BBCC03].

4.1.3 Reference architecture - the horizontal view

The components of the reference architecture reflect the important interoperability aspects from an architectural perspective. In the center of this horizontal perspective are the references between processes and services [BBCC03].

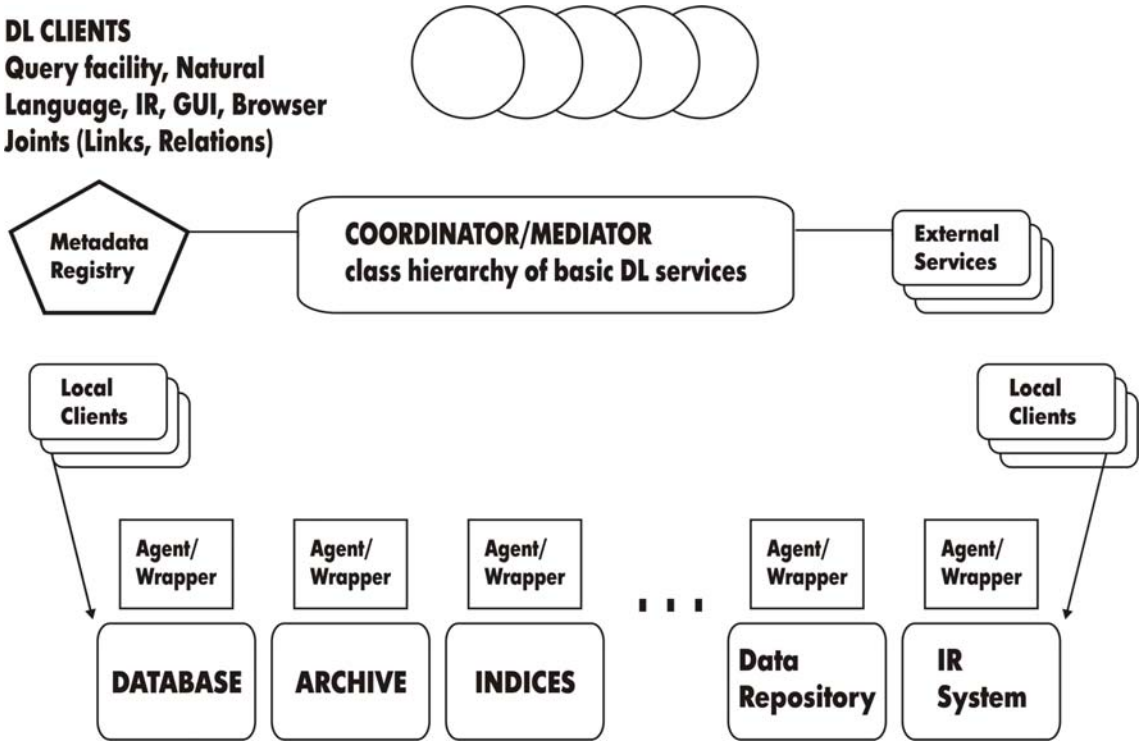

Figure 3 "Interoperability architecture from the horizontal perspective". Adapted from [BBCC03].

In the figure 3 above we find the depiction of the reference architecture. Heterogeneous sources and services are combined into several digital library services. Those are interoperable and together form the global interoperability layer for digital library systems. The original data sources and services can be accessed through specific access methods, which are defined by agents or wrappers. The latter also transform the local data models into the global information model [BBCC03]. The original data sources and services, though combined into few general ones, are autonomous so that local clients can access these servers [BBCC03]. The mediator layer contains the basic digital library services, but external services also can be integrated. The mediator layer enables the usage of advanced query languages by clients. It is important that the query facilities include the capability to combine information from different data sources [BBCC03].

Reference architecture gives a different view at the topic of interoperability than the interoperability layers perspective, and together they form a sufficiently full depiction of the interoperability components structure.

4.2 Descriptions of objects and repositories

The research issues at the descriptions level deal with the techniques and resources needed to request objects from digital collections correctly [Scha95]. These resources are mainly situated at the semantic level. They are closely related to the mechanisms for describing the meaning and purpose of the objects in the collections. Such mechanisms are: metadata at the semantics-level, which describes the value and meaning of the object substructure; computed descriptions, which extract meaning deduced from object content [Scha95]. Other mechanisms are “unification - merging the semantics of the metadata across descriptions” [Scha95]; and “organization - clustering the descriptions to facilitate navigation” [Scha95].

The description issue is very important, as most of the information searching systems use descriptive metadata. This metadata describes the data in library catalogs, and is often produces by abstracting and indexing services. The creation of metadata by humans takes a long time in regard to the huge amount of information in digital libraries and it is costly as well. Alternative approaches must be found. One of them is to oblige the creator of a digital object to produce small amounts of descriptive metadata for each created piece of information [Arms99]. Another approach is provision of metadata by the administration of a digital library. This metadata can be created either manually or automatically by a program or tool. The metadata can then be supplied to an automatic indexing program [Arms99].

In order to provide a rational view of collections of digital objects, they must be described in a consistent manner which can facilitate the use of mechanisms such as protocols that support distributed search and retrieval from dissimilar sources [LyGM95].

The research topics in the field of descriptions include according to Lynch and Garcia-Molina the “definition and use of metadata, the use of computed descriptions of objects, federation and integration of heterogeneous repositories, clustering and automatic hierarchical organization of information,

and algorithms for automatic rating, ranking, and evaluation of information quality, genre, and other properties” [LyGM95]. Other important issues involved are knowledge representation and exchange, and the definition and exchange of ontology for information context [LyGM95]. Research in description of objects and collections of objects provides the foundation for effective interoperability.

4.3 Users

The working environment of a typical user consists of many parts, such as electronic mail, word processing, and programs, applications and databases specific to the individual's area of work. In addition, the working environment usually includes a lot of information not in digital form, such as books, papers, or video tapes. Collaborative digital libraries are just part of this working environment. The capability of users to interact and work with digital objects, to adapt and process them, and to add them to the personal collections is an important area of research [Arms99].

User interaction and user interfaces are next important research subjects. This topic is rather complicated and according to Arms [Arms99], the success of research in this area was expected to depend mainly on intuition and creativity, rather than on a technological progress. These considerations didn’t meet the reality, though. An example of successful combination of creativity and technological research is the development of web browsers [Arms99].

Digital library collaboration can only exist with effective possibilities of communication between its participants. User access to digital libraries brings an additional layer of complexity. The management of user access is an area of digital library infrastructure where legal, policy, and technical issues most strongly interact [FMLM03]. Therefore, an important research issue is the one dealing with the interaction required for users to adequately access a digital collection. This research topic deals with, firstly, user needs and their satisfaction - user assessment, user interface, etc. and, secondly, with the user contributions - enabling the users to organize the digital collections themselves

for better personal access [Scha95]. Users should be able to define their preferred access context as a combination of their historical access patterns, their current interests, and their personal level of knowledge [FMLM03]. The management of user access is the area of digital library infrastructure where legal, policy, and technical issues most strongly interact. The resolution of possible conflicts requires policy decisions for use of the digital library infrastructure by its users [FMLM03].

The usability of computers by people is the subject of intensive research in such fields. The research is relevant for the theoretical understanding of the interaction of people with computers. Based on this information, we can develop models for design of appropriate computer systems. [Arms99].

While regarding a computer system from the user point of view, the first intention is to concentrate on the design of the interface between the user and the computer, but this would not be correct, as usage comfort depends of the total system [Arms99]. All the components must work together efficiently to create a successful and convenient digital library for all participating sides [Arms99].

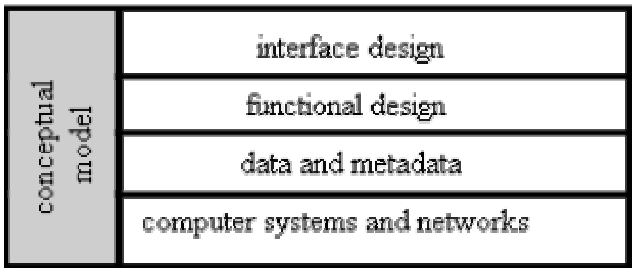

Figure 4 “Aspects of usability”. Adapted from [Arms99].

The figure 4 shows an example of the usability components and the design of user interfaces. The user interface is always built on a theoretical model that describes the way in which the system is used. There are several basic conceptual models that are used in the design of digital libraries, and there are models that are used for the collaboration of digital libraries. In practice, most digital libraries combine concepts from different models [Arms99].

The right side of the figure 4 shows the layers that are needed to implement any conceptual model. The top layer is the interface design, the presentation on the

screen and the actual handling by the user [Arms99]. This layer comprises the considerations of menus, buttons, links, fonts, colors, etc. The next level is the functional design, which characterizes the offered functions, such as selection of fractions of a digital object, searching a list or sorting the results, information about the functions, and the manipulation of objects that have been provided on a screen [Arms99]. These functions are enabled by the two bottom layer: the data and metadata provided by the digital library, and the underlying computer systems and networks. [Arms99].

The issues of user interfaces and human-computer interaction are an extensive field of research. In the field of digital library collaboration there are some specific problems that are central to the progress of the area. The key areas according to Lynch and Garcia-Molina are display of information, visualization and navigation of large information collections, and linkages to information manipulation/analysis tools. The need for a more comprising understanding of user wishes, aims, and activities while making use of digital libraries is the starting point for designing effective systems. Digital library systems must become far more effective in adapting to differences in the capacities of personal computers and network connections and in presenting appropriate user interfaces [LyGM95].

4.4 Collections

Digital libraries consist of collections, also in the form of distributed depositories [FMLM03]. Collection management is a research topic, which has been neglected and didn’t receive attention that it deserved from the start. Traditional libraries have developed their methods for collection management, where huge collections of material were managed by comparatively small teams of people [Arms99]. At the same time, digital libraries were significantly labor intensive in the beginning. In the enthusiasm of creating digital collections, the necessity of organizing and preserving the information over long periods of time, especially in the field of collaborative digital libraries, was neglected [Arms99]. Conversion

of physical materials into digital formats illustrates the difficulties of collection management. The best way to transfer gigantic collections to digital format so that the cost-and-quality relation would suit has to be found. The usefulness of the today's efforts in the long term has to be taken into consideration [Arms99].

In the opinion of Lynch and Garcia-Molina [LyGM95], collection management and organization research is the area where traditional library missions and practices are reinterpreted for the digital library environment. Progress in this area is essential if digital library collections are to meet successfully the needs of their user communities [LyGM95].

An important research field is the collections management needed for their effective organization. According to Schatz [Scha95] here we have several topics: archiving, preservation and insuring that access to the digital collection is possible on a "permanent" basis; virtual collection development dealing with providing tools to construct/organize a distributed collection; and repository management - tools to update and maintain a digital collection [Scha95].

The central problems in the management of digital collections according to Lynch and Garcia-Molina are the policies and methods for incorporating information resources on the network into managed collections, rights management, payment, and control issues. The authority and quality of collection content in digital libraries is of central concern to the user community [LyGM95].

The organization of large collections of online materials is a complex issue. Many of the problems are common to the materials of an electronic journal, a large web site, a collaborative digital library or a compound information service. The issues include questions about loading information in changeable formats, and organizing it for storage and retrieval [Arms99]. For access around the world, several copies are needed, using various techniques of replication. The problems are enlarged by the fact that digital information changes, and it changes continually. It is difficult to follow the small differences and whole collections have to be reorganized at unpleasantly frequent intervals [Arms99]. Many of the research topics that are essential for interoperability between collections are just as important for organizing large collections. Especially

current research on identifiers, metadata, and verification applies to both collection management and interoperation among collections [Arms99].

The long-term preservation of digital materials also proved to be one of the most relevant research topics in collection management. Physical materials, such as printed books, have the quality to be readable over decades, even if they were neglected during a long time [Arms99]. Digital materials are the opposite. The media for data storage has quite short life expectation. The data has to be refreshed by explicit actions like copying periodically onto new media. Even if the data is preserved, the problems remain, as the formats in which information is stored are often replaced by new versions. Formats for text processing and image storage will be obsolete and hard to use in less than ten years. There is an acute need of solutions for making collections of materials readable and useful in the future [Arms99].

All these topics are now being recognized as complicated and crucially important.

4.5 Legal issues

The legal issues of digital library collaboration comprise the matters of the material sharing, the control over the ownership, the privacy management, the material and reuse etc [LyGM95]. The EU-NSF Working Group on Intellectual Property and Economic Issues [NWDH96] outlined, that a key problem in digital publishing is that there are many gray areas in the existing legislation concerning the details of computer-based network systems. In some severe cases, strict application of the current law can even result in rigorous restrictions that eradicate advantages made possible through technology [NWDH96]. One possible way to improve the uncertainty about current copyright and intellectual property laws is to rely stronger on contracts and license agreements [NWDH96]. But this cannot eradicate the need for adapted legislation.

Many legal problems result from the international scope of digital libraries and international origin of the collaborative partners. This way, communication and information exchange can take place across many governments and legislations and it is not obvious which legislation must control what content and activities [NWDH96].

A significant security issue related to provision of content through networks is rights management. Unauthorized use and reproduction of information is a key concern to content owners. Substantial research in this area is needed and technical means have to be developed to discover and prevent such unauthorized use [Herz98]. There are several research issues relating to the legislative side of the interchange of the digital content. All of these issues are important for the collaborative digital libraries as well, because they carry the responsibility not only for themselves, but also partially for their collaborative partners. Collaboration then has to be established in a way that would ensure that all of the actions are carried out in accordance with the legislation. Which legislation should be applied in case of international collaboration is another problem.

The “The EU-NSF Working Group on Intellectual Property and Economic Issues” at [NWDH96] described some of legacy research issues in the classification you will find below.

4.5.1 Cached Copies

The copyright legislation generally forbids the replication and distribution of a piece of information without having acquired a license from its owner [NWDH96]. Besides, copyright protects information from being plagiarized, even if there is no copy made, this way the copyright serves as a major protector of intellectual property. These both protection modes are absolutely crucial for the protection of digital works, but the notion of a copy has to be changed or refined for digital objects [NWDH96].

In the digital world it is inevitable to download copies of digital objects from a server into the local memory of a computer before they are displayed to the user [NWDH96]. The process is called caching and it is indispensable for achieving satisfactory performance [NWDH96]. Caching at least parts of an object into local memory cannot be evaded [NWDH96]. At this point, the following questions arise: is a “cash copy” a real copy in the legislative sense? If not, then what should the lifetime of cached data be, or how long can a cached document stay on a computer before it turns to a "true" copy in the legislative sense? [NWDH96]. It is clearly not desired by the creators of a digital object, that the cashed data is to be destroyed eagerly, perhaps while the user is still using the data. This cannot be the purpose of the corresponding legislative act. Yet it is also unacceptable that the cached information is kept for ever [NWDH96]. One proposed solution is to keep cached objects in the local memory of computers for the duration of the connection with the server, which provided access to the document, or to let cached data bear a lifetime value that would specify when the data is to be destroyed [NWDH96]. To solve this problem, an explicit and practical distinction between “true” and cached copies of digital data is to be defined [NWDH96]. The true copies would be then the subject to copyright and the cashed copies would be seen only as temporary data that is produced for efficiency purposes.

It could also be important to differentiate between voluntary and involuntary acts of copying [NWDH96]. A voluntary copy would be the one made by explicit act of copying initiated by a user. An involuntary copy would be the one automatically produced by a program, without the user taking notice of it. The legislation would also then have to distinguish between these two cases [NWDH96].

This topic is also very important for the collaboration of electronic libraries, because they are usually able to use the resources of each other. It would require strict coordination to figure out, where what copies are stored, and during what period of time.

4.5.2 Access and referencing

The notion of distribution is widely applied in the copyright law, but for the electronic publishing organization it might be less significant than the notion of access. The physical data cannot be viewed or used outside the place where it is being kept unless it is physically copied and distributed to other locations. Opposite to this, digital objects stored in on-line digital libraries can be viewed over a network without have to be copied [NWDH96]. For public digital libraries connected to the World Wide Web and open to common access, this could have as a result that the users have no interest in possessing a copy of a published object, because these can be accessed at any point in time and from any location with no restriction [NWDH96]. At the moment, an important technique to oppose unauthorized use and distribution is to distribute data encrypted within a secure container, which is opened and used only by trusted software of hardware module [Herz98], but it is not a suitable solution to the challenges of the World Wide Web.

Direct access can result in subtle legal problems. By bringing a copyright protected data on an unprotected public server, one enables an act of copying and distribution by others, although one does not directly violate copyright. Here, although the object in question is being protected by the copyright, enabling online access to it makes it available to the open public [NWDH96]. Access may even occur indirectly, due to the WWW technology supporting hyperlinks. In this case, one may be violating the law simply by publishing a document that has links (hypertext references) to other digital objects protected by copyright and kept in distant web collections [NWDH96].

A somewhat different case would occur, if an author includes a link to data that is originally not restricted by copyright, but becomes copyrighted later without the author taking notice [NWDH96]. Here, the following question arises: who is responsible if the protected document is copied or accessed through that particular link? A similar situation occurs if the contents of the linked data change unexpectedly, e.g. providing illegal material, also without the author taking notice [NWDH96].

Another example: supposing that the creator of a digital collection receives funding through advertisements on the main page of the web-site of his collection. Then, if someone generates links enabling users to access digital object of this collection directly rather than through the main page represents a likely financial danger, because the hit count of the main page decreases and companies placing advertisement will no longer wish to use this site for their promotion [NWDH96].

All this stresses the importance of access control. It is very important to come up with suitable legislation for regulating all these concerns. Most important, the legislation must be appropriate enough not to unnecessarily restrict the enormous potential of collaborative networked digital collections [NWDH96].

4.5.3 Virtual Spaces and Non-human Agents

The problem here is that in computer-based systems, programs execute actions, seemingly initiated by users, but this can be both with and without them knowing about it [NWDH96]. A program can also pass on one or more tasks to other programs, possibly located on different computers. Those, in turn may delegate the tasks to other software components. This way, a single user request might result in a highly complex interaction. So, a common reader is acting as a user, who may initiate a trigger process of additional actions on other objects. This functioning is totally different from what we are used to in the field of physical objects like printed documents. Of particular importance are acts of indirect infringement or violation of copyright and other legislative regulations. The problem is that it is not always obvious whether the violation can be attributed to the user who started the action or to some other participant, who facilitated the activation of an intermediate program [NWDH96].

4.5.4 Privacy of User Profiles

Traditional libraries have always safeguarded the privacy of user records quite eagerly. There are laws in many counties all over the world that prohibit libraries to disclose information tracking user activities in the library and oblige them to protect the user privacy. This way, another significant research issue is to find out to what degree the developers of digital libraries are required to consider the aspects of user privacy and develop features for protection of anonymity [NWDH96].

For example, it would be easy to pile up a list of information that particular readers might be interested in and to what price. This would be precious information. It could be used to regulate the prices of products, maybe in favor of the customer, but maybe not. This information could be also used to adjust the performance factors of the whole underlying system [NWDH96].

Furthermore, the digital library developers might not just compile such user profiles themselves for their own use, but also to offer user profiles for sale or exchange with other digital libraries. Here, the question is, whether this would be fair use of user data? Would it be legal to put a person's name on a list that is published without the persons' consent and make profit by that? [NWDH96]. All of these questions need to be investigated and solved as soon as possible.

4.6 Economic issues

The continuous alteration of technology nowadays leads not only to the transformation of the information form from physical to digital, but also to changes of economic models being applied to it. Rather than using specific equipment and labor services for printing text or developing film negatives, information can be made accessible and usable in other, non-traditional ways [NWDH96]. A digital library is one of the most active participants of the new

information access models. Nowadays, it operates as an institutional link between readers, authors, and other digital libraries [NWDH96], by holding large collections of digital information and by enabling the readers to access their collections for free or for a favorable fee. So, in order to stay competitive they have to develop alternative economic models [NWDH96].

The revolution in digital publishing changed the world of information and unsettled historical balances existing for a long time and the system for traditional cooperation forms between publishers and libraries. This change offered significant advantages, but also caused many problems. The main advantage is that information suppliers can bring up information faster, cheaper, and for a greater number of user than ever before. The users, in turn, receive direct access to a large amount of information materials situated all over the world [NWDH96]. The disadvantage is the rise of major intellectual property and economic concern issues.

Digital information can be copied very fast, at almost no expense, and without any loss of quality. Due to this, the publishing of digital information bears many risks both for publishers and for the digital libraries [NWDH96]. Electronic publishing changes the way institutions produce and distribute intellectual work. These new form of information distribution could bring new forms of business into life. In this new market for information, it is very relevant to resolve the question of how digital libraries can function in order to be economically viable or even profitable [NWDH96].

Agreeing on a certain financial model is difficult, because most of the information or information services offered today are offered for free [NWDH96]. Although most of the services or materials offered were implemented at a substantial cost, we are able to use them for free, and so this emerges to a standard that has to be kept [NWDH96]. Another unacceptable situation could arise, if data would be extracted out of a public web site in order to be republished or used for publishing of value-added information. This action could seriously interfere with the commercial interests of the collection owner [NWDH96].

4.6.1 Operating costs

Although the expenses of digital publishing and distribution are considerably less than those of printing, they still exist and have to be considered. There are tasks such as information editing, filtering, refining, converting and preserving, and although the way of creating and distributing of information has changed, such tasks still require substantial resources. A librarian still has to gather material that can arrive in any possible digital format, structure the collection etc. All of this consumes time and funds. Furthermore, there are new tools and technologies for processing and presenting information being continuously produced, and this leads to the necessity of permanent training of the editorial staff, so that it can be always well prepared to take full advantage of the new possibilities [NWDH96].

An important cost factor is the price of indexing and archiving [NWDH96]. New technology offers enormous potential, but it is not enough for a work to be created or acquired. Every piece of information still has to be classified, catalogued, and stored in a database from where it can be retrieved on request. These tasks are being completed by the new software systems support at little incremental cost, but this does not mean that they are free. One reason for the ever-lasing expenses is the rapid changes of modern technology and media leading to the changes in forms and formats of digital information. This leads to the necessity of periodical upgrading and extending of such systems, and this is also rather expensive both in terms of time and money [NWDH96]. Considering the fact that the users are used to free access to the great part of the digital material, new economic models are necessary that can account for the costs mentioned above and still offer enough space for future development [NWDH96].

4.6.2 Revenue model

As mentioned above, the cost structure of copying and distribution of digital information differs from that of physical information. It has changed with the development of digital technology. Copying of conventional information was expensive, besides, there was a definite price-quality relationship: it is less expensive to make a copy yourself than to buy a book, at the same time a book will always be of a better quality that a self-made copy.

The situation with digital information is totally different: it costs nothing to make a copy of a piece of information, and the quality is just as good [NWDH96]. We have come to a point, where the suppliers of information are forced by the rising competition to offer their products and services at the replication cost, which is zero [NWDH96]. These new circumstances lead to the necessity of a serious modification of the traditional pricing models. There is a strong need for new forms of economic relationships in this area, allowing co-existence of competition and profit. These new economic models must differ from the customary pricing approach. Finding out the production cost of one additional unit of a product is of no interest anymore. Instead, new cost factors arise, like constructing, maintaining and managing a collection as a whole. These new expenses set up new monetary approaches for ensuring long-time existence and maybe profitability of digital libraries [NWDH96].

The scenario, where the incremental production and distribution cost of each further unit equals to zero or almost zero, is quite new and different from what we are all used to. As noted in [NWDH96], a good starting point for the development of new models could be the nonlinear pricing schemes. This means that the revenue expected from a certain quantity of products does not stand in a linear proportion to that quantity. The idea is that information publishers might be able to profit more by requiring not a linear but a variable price per unit. This approach is expected to be suiting in respect to dissimilar purchasing power of the clients [NWDH96].

The need for new pricing schemes is strong, but there is vagueness concerning the product bundling structures for the cost and price determination [NWDH96].

It is not easy to define those parameters that should serve as a basis for product pricing. In a digital collection we deal with items as, for example, journals. In order to determine the price of a journal, we have to consider that it consists of multiple objects: a journal comes in volumes, which in turn contain many articles. An article also consists of many objects: header, text sections, figures, pictures, video etc. These are all dimensions, along which diverse bundles could be constructed. Besides, as mentioned above, nonlinear pricing can be applied to some these objects [NWDH96].

It is really difficult to optimize all of these factors, and the more complex a pricing scheme would be, the less consumers would be satisfied with it. This way, the most thorough optimizing could even end in the decrease of the product quantity sold [NWDH96]. The optimal dimensions of a pricing model will depend on the defined heterogeneity of user preferences in regard to different dimensions, and the costs of production and distribution of bundled objects. Since user preferences are unstable and can change any time, it will be essential for institutions to remain flexible in their pricing. The publishers will have to follow the demands of the market and to adjust their pricing models continuously [NWDH96].

4.6.3 Market positioning

In the world of physical information, traditional libraries never really played the role of rivals in the market of information [NWDH96]. With the development of digital information, information providers have to face the necessity of becoming players in the global market for digital information. The viability of a digital library depends on its popularity and market position in the World Wide Web [NWDH96].

Competition in the World Wide Web is especially hard [NWDH96]. The popularity of a digital library can rise very fast, but it can fall just as fast as well. Due to the existence of the e-mail and numerous interest groups, the reputation, either good or bad, spreads around very fast, possibly within a few hours. And it

is much easier to lose a reputation, than to build a new one [NWDH96]. The users know what they want; ant this is information and services of high quality. Such data being offered can be sold for a reasonable price or funded from other sources, presuming the library is popular, but low-quality products will make the users stay away from a digital library, even if these products are offered free of charge.

The technical characteristics of an Internet site are maybe not less important than its content. Users often choose a particular site just because it responds faster to their requests. Even if they know a site containing information of a higher quality, they would still prefer a site that is working faster [NWDH96].

In the case of a website that does not satisfy the users' needs, it will be very hard to offset even the lowest operating costs. So, the survival of a digital library will depend on its information service quality [NWDH96].

More competitive in this market situation will be the collaborative digital libraries, due to the larger amount of information and services offered. Recovering costs will be also easier in this case, due to the synergy effects and other benefits we mentioned above.

Traditional public and academic libraries also find themselves in a totally new situation. Unlike earlier, now authors and publishers of articles, books and journals have the possibility to provide direct access to their works. This leads to a much weaker position of traditional libraries, a position which has to be newly identified. The traditional libraries seem to have other choice but to enter the free trade market of digital information and find their market position. Traditional libraries have to start offering their information materials in a digital form online. Some information, like that of cultural libraries and museums will have to be digitized and made available to the wide public in this new form. Other libraries, like scientific and academic libraries, can start to offer scientific journals and reports, or materials for distance-learning [NWDH96].

4.6.4 Market and legal incentives

As we already mentioned, one of the most important research issues in digital publishing is the development of methods and incentives to avoid illegal duplication of digital documents. One way to solve this problem is to develop legal mechanisms that would keep away users from copying a digital object. The thought is that a possible violator of a property right would compare the benefits from use of an object without paying the provider and the expected costs of penalties for the case that he will be caught. But digital material can be accessed and replicated easily and mostly without a trace, and so this legislation might not bring the desired results [NWDH96].

A different approach to reach the same goal could be the use of market incentives. Such an approach leads to the situation, where it would more expensive to try to use a product illegally rather that to buy it. An example of such an incentive are copy-protection schemes. Here, a company would distribute digital objects in form of "protected" packages that cannot be replicated directly. Of course, it is possible to crack the protection, but the idea is that it would cost more to do this rather than to buy a product legally. [NWDH96].

In reality, the legal and the market methods are often combined to achieve best results. So, in the world of physical objects, a violator is to be discouraged from illegal copying is discouraged both by the anticipated legal penalties and by the cost of photocopying. Since the cost of replication of digital information is due zero to the rapid development of digital technologies, the balance between the traditional methods of preventing infringements shifts, and new incentives have to be found [NWDH96]. Collaborative libraries have to find common solutions for protection of their collections, rather that trying to fight law violation alone.

4.6.5 Funding sources

As stated by William Arms [Arms99], the difference between digital information and traditional information is that the costs are almost or just the same no matter if the materials are being used by somebody or not [Arms99]. For both digital and electronic publications, such tasks of creating digital information like composing, reviewing, editing, designing and formatting are the same [Arms99]. The costs of distributing digital materials are tiny and so the total costs consist only of the production costs, where as the costs of physical publishing consist of both production and distribution costs. So, costs of digital objects do not depend on the number of objects produced. Saying it in economic terms, costs are fix, and the marginal cost is near zero [Arms99]. As a result, only the production costs have to be covered by the revenue from sales of digital products. Once the revenue from the number of items sold covers the costs of creation, all further sales are pure profit. Unless this break-even point is reached, the product is destined to make a loss [Arms99].

All organizations large or small, commercial or non-profit have to cover their costs somehow [Arms99]. Here, several ways of funding exist. The various approaches can be divided into models, in which the user or the digital library pays for access to the collections, and, on the other side, open access, where the funds come from the creator or producer of the information [Arms99].

The fact that there are the tremendous amounts of excellent material on the web that are available for open access, with no requirement for payment by the reader is extraordinary [Arms99]. This comes from the aspect that many writers and creators of information are keen that their materials should be seen and are ready to cover the costs themselves. They are willing to invest their own resources to make their materials openly available. Such creators are for example “researchers seeking for professional recognition, government agencies informing the public, all types of marketing, hobbyists, and other recreational groups” [Arms99].

When creators prefer the open access model because they wish to have their work widely used, there are still costs that have to be funded somehow

[Arms99]. One important source of funds is grant funding, which has now become institutionalized [Arms99]. Grants are usually short-term, but they can be renewed [Arms99].

Another funding possibility, e.g. applied by many web search firms, is providing open access that is paid for by advertising [Arms99]. This is the model that is used also by broadcast television in many countries [Arms99]. So, the user does not pay directly for information. The only thing that is expected from him is his attention to advertised products, and his time to follow the links leading to them. The revenues come from advertisers. Most of the companies using this model are now profitable [Arms99].

Research teams are especially interested in having their work widely read [Arms99]. They cover the costs of open access materials and the costs of their scientific research from the same source. Research is mostly funded by government or by large companies, and this is where the money for open information on the web comes from.

Many open access collections are created and maintained by government departments [Arms99]. They provide a lot of information that is important for their nations, such as the hurricane tracking service provided by the United States [Arms99], or information about the social issues etc. Most of the information they provide is important only in the short-term, but many of their collections are of long-term value [Arms99].

Many open access library collections come from private individuals [Arms99]. Some of collections contain high-quality information, e.g. devoted to sports and hobbies, fan clubs, and privately published poetry or novels [Arms99]. These collections are just funded by their producers [Arms99].

5 Technical support for digital library collaboration

The ultimate value of digital resources to the user will depend not only upon the quality of the content and the organization, but also upon the data management systems. The digital library interoperability must be therefore supported by sufficient and high professional IT means [Pede00]. This support is usually implemented at the last phase of digital library collaboration establishment. This stage has to be carefully planned and discussed, as any mistake or misunderstanding will cause high expenses. The collaborating libraries have to agree upon the hardware and software as well as upon standards, tools and languages to be used jointly [Pede00]. This process is difficult as there are so many offers to choose from. The range is wide and it changes rapidly from year to year. Tools and standards have to be popular and comfortable enough, but they should allow expansion and development [Pede00]. The main problem is that different institutions may use different paradigms for arranging their products and implementing internal and external access control [NWDH96]. This means that appropriate standards and conversion mechanisms must be developed not only to pass control information from one system to another, but also to achieve semantic interoperability for the various control levels [NWDH96]. So, there is a strong need for establishing common terms and conditions that can accommodate the enormous potential complexity of the collaborative processes, and languages and tools for understanding, negotiating, and contracting [NWDH96].

In the rapidly changing IT world products offered are so numerous and are changing so fast that it would be useless to give a full classification and a broad overlook of all of them. Therefore, we concentrate on the most popular and relevant to the field of digital library collaboration standards and technologies.

5.1 Dublin Core Metadata

Metadata means data about other data. It is the structured data about digital objects and collections. This term usually refers to “descriptive information about Web resources” [HILL01], information specifying such data as the title of a document, the author of a resource, the date of its creation etc. Metadata according to Hillmann [HILL01] is a very important topic today since it allows a unified and standardized description of the tremendous amount of the existing information resources. Metadata is the key to locating, using, and preserving digital content [HILL01].

The Dublin Core Metadata is an important standard for the digital library collaboration because it provides the possibility of standardized description of the existing enormous amount of digital information in the digital libraries. It is significant, because it gives the libraries the unified possibility of browsing, searching and retrieval of information being stored in all digital libraries standing in collaboration. This standard is a quite easy to implement solution to the problem of a unitary data description and it is compatible with other new technologies and standards like HTML, XML, RDF or Open Archives Initiative Protocol for Metadata Harvesting.

A metadata record consists of a set of attributes, or elements, necessary to describe the resource, like creator, usage rights, type of document etc [HILL01].

There are three different types of metadata, all of them important for the usability and preservation of the collection over time [HILL01]. The first metadata type is the descriptive metadata. This is the type the search tools are dealing with, as it provides information that allows discovery of collections or objects by search tools, and it provides sufficient context for the user to understand what the search results are. The second type of metadata is structural metadata, which describes the associations within or among related individual information objects – one object contains another etc. [HILL01]. The

third type is the administrative metadata that facilitates access, management, and preservation of the digital resource [HILL01]. It can describe the tools necessary to access the object, or attributes such as image resolution and file size, as well as rights management information.

Much of the utility and long-term maintenance of digital collections is dependent upon the types and quality of metadata that is used within the collection for the digital object description, but agreed-upon standards for metadata are only beginning to emerge [HILL01]. Without agreed-upon metadata standards all the bigger digital collections would be useless.

The Dublin Core addresses the need for "standardized descriptive metadata" [HILL01]. The Dublin Core metadata standard is a simple and understandable but effective and comfortable for use set of 15 elements for describing a wide spread of types of resources [HILL01]. Each element is optional and repeatable. Metadata elements may appear in any order. The 15 elements are [HILL01]:

- Content related: “Coverage”, “Description”, “Type”, “Relation”, “Source”, “Subject” and “Title”.
- Related to intellectual property: “Contributor”, “Creator”, “Publisher” and “Rights”.
- Related to instantiation: “Date”, “Format”, Identifier” and “Language”.

Dublin Core can be represented in several different syntaxes, including: HTML, RDF/XML and the generic form (Element="value") [HILL01].

An example of elements in the generic form for this work could look the following way [HILL01]:

Title=”Collaboration of digital libraries”

Creator=”Maria Shustitskiy”

Subject=”Digital library collaboration”

Type=”Text”

Identifier=”http://wwwai.wu-wien.ac.at/hahsler/dlibcollaboration.doc” (This is not a real path, but an example of a possible “Identifier”)

Date=”2004-30-11”

Language=en

Type=”Text file, 1.5MB”

For encoding this metadata using HTML 4.0, we must indicate the Dublin Core elements by using the prefix “DC” [Kunz99]. Encoding the element “Title” into the META tag of HTML would be possible this way:

```
<meta name   =  “DC.Title”
        content = “Collaboration of digital libraries”>
```

A complete HTML file with the embedded description using Dublin Core metadata could be:

```
<html>
 <head>
 <title> Collaboration of Digital Libraries</title>
  <meta name    = "DC.Title"
      content = " Collaboration of digital libraries ">
 <meta name    = "DC.Creator"
      content = " Maria Shustitskiy ">
 <meta name    = "DC.Type"
      content = "Text">
 <meta name    = "DC.Date"
      content = "2004-30-11">
 <meta name    = "DC.Language"
      content = "en">
 </head>
 <body>
       Digital libraries are organized collections of digital information…
 </body>
 </html>
```

(Adopted from [Kunz99]).

Another possibility to apply Dublin Core elements is to embed them in a scheme of Resource Description Framework (RDF). RDF is a framework that can be used for numerous metadata schemes and that can both be processed by machines and read by people [Hill01]. XML (EXtensible Markup Language) is used by RDF to define structure and semantics. This is a decentralized approach allowing building of schemes for required in different cases. Here is an example of a metadata schema in an RDF/XML framework:

```
<rdf:RDF       xmlns:rdf=http://www.w3.org/1999/02/22-rdf-syntax-ns#

                xmlns:dc=http://purl.org/dc/elements/1.1/>

  <rdf:Description

rdf:about=" http://wwwai.wu-wien.ac.at/hahsler/dlibcollaboration.doc">

        <dc:creator>Maria Shustitskiy</dc:creator>

        <dc:title>Collaboration of digital libraries</dc:title>

        <dc:date>2004-11-30<dc:date>

  </rdf:Description>

</rdf>RDF>
```

(Adopted from [Hill01]).

There are two types of linkage between the metadata and the document: the metadata can be embedded in the resource - this approach is taken most often with documents encoded using HTML - or in a separate file. According to Hillmann [HILL01], the Dublin Core standard does not prescribe either type of linkage, leaving the decision to each particular implementation. Simple tools have been developed to make provision of Dublin Core metadata within HTML encoded pages quite easy. One such tool, DC.dot, extracts metadata information from an HTML document and formats it so that it can be edited, then cuts out and pastes back into the HTML header of the original document. The Dublin Core favors document-like objects (because traditional text

resources are fairly well understood), but it can be applied to other resources as well [HILL01].

The benefits of Dublin Core are simplicity of creation and maintenance, commonly understood semantics, international scope, availability in several languages and extensibility [HILL01].

5.2 Open Archives Initiative Protocol for Metadata Harvesting

The Open Archives Initiative Protocol for Metadata Harvesting (the OAI-PMH) is a development of the Open Archives Initiative. It provides an “application-independent interoperability framework based on metadata harvesting” [Bree02], which is applied to support the creation of value-added services information for the users of the information [Bree02].

The protocol is a development important for digital library collaboration, as it is enables the providing of value-added services to the users of digital information. Such services could be implemented by the collaborative digital libraries, and they can be used by them as well for a better employment of collaborative benefits.

The OAI-framework is based on information repositories, or "data providers" [Bree02] that use a given set of protocols to make their metadata available to "service providers" [Bree02] that create new information resources. OAI-based services can aggregate the metadata of many OAI repositories. OAI does not operate with complete works of digital content, but only works with metadata. Usually the metadata includes links to the source of the original information repositories for access to the digital information [Bree02].

According to Marshall Breeding [Bree02], there are two classes of participants in the OAI-PMH framework. First, there are data providers, which administer

systems that support the OAI-PMH as a means of exposing metadata. Second, there are service providers, which use metadata harvested via the OAI-PMH as a basis for building value-added services [Bree02]. A harvester is a client application that issues OAI-PMH requests [Bree02]. A harvester is operated by a service provider as a means of collecting metadata from repositories. A repository is a network accessible server that can process the six OAI-PMH requests. A repository is managed by a data provider to expose metadata to harvesters.

The six OAI-PMH requests are the following:

- GetRecord. This request is used to retrieve a single metadata record from a repository [Bree02]. Required arguments specify the identifier of the item from which the record is requested and the format of the metadata that should be included in the record [Bree02].
- Identify. This request is used to retrieve information about a repository. Some of the information returned is required as part of the OAI-PMH [Bree02]. Repositories may also use this request to return additional descriptive information [Bree02].
- ListIdentifiers. This request is an abbreviated form of ListRecords, it retrieves only headers rather than records [Bree02].
- ListMetadataFormats. This verb is used for retrieving the available metadata formats from a repository [Bree02]. An optional argument restricts the request to the formats available for a specific item [Bree02].
- ListRecords. This verb is used to harvest records from a repository [Bree02].
- ListSets. This verb is used to retrieve the set structure of a repository, useful for selective harvesting [Bree02].

The term "archives" refers to any kind of digital content repository [Bree02]. The OAI metadata-harvesting protocol operates behind the scenes only between data providers and service providers. No special software is needed to search

an OAI-based information service. The users need not even be aware that OAI was used to collect the metadata [Bree02].

OAI is an open framework. Any public, private, or commercial entity can build systems that comply with this protocol [Bree02]. The initiative makes no distinctions with information access. An OAI service provider can offer free and universal access, restrict access to specific communities, or impose fees for access. OAI takes a neutral stand on access control and business models [Bree02].

Below you will find an illustration of the functionality of the harvesting based on OAI-PMH. As you see, it can be as well combined with searching based on, e.g. Z39.50.

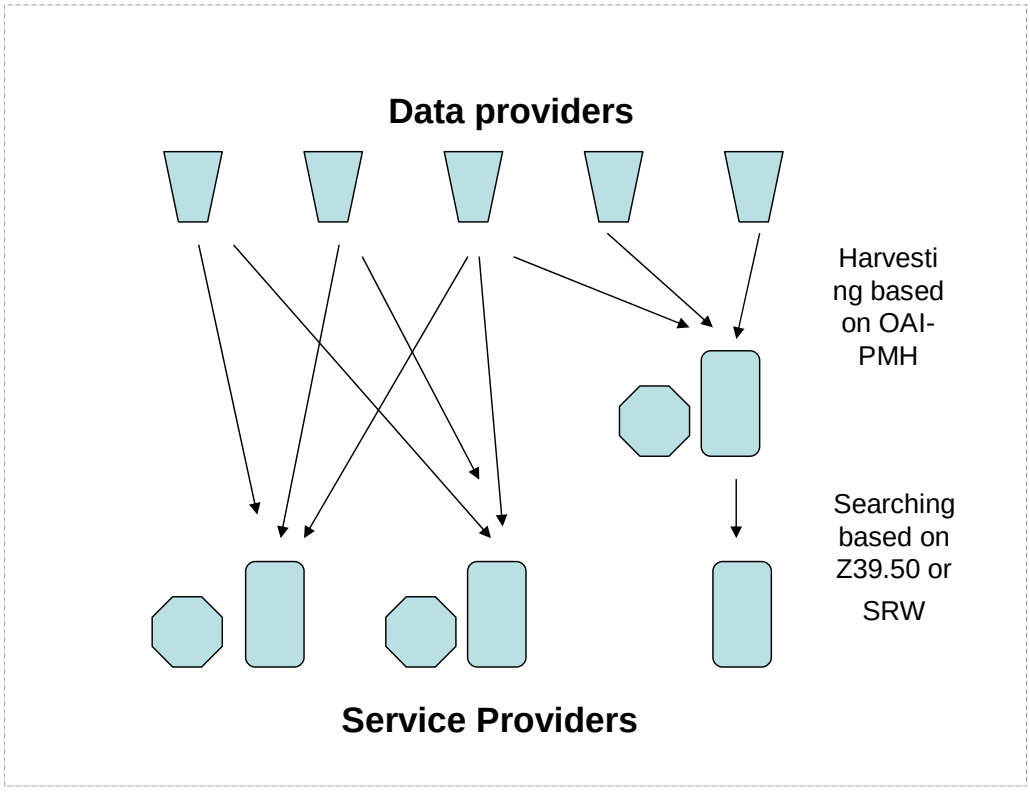

Figure 5 “Harvesting combined with searching”. Adapted from [Bree02].

The OAI protocols were designed to be very simple and efficient. It only takes a few days of programming for enabling an existing information repository to function as an OAI-compliant data provider. Building services is usually a little more difficult [Bree02].

5.3 Z39.50

Pederson [Pede00] describes Z39.50 as a protocol that defines how to search a remote catalogue. This is a standard the National Information Standards Organization (NISO) is responsible for. This protocol defines “how to formulate the query and how to present the result of the search generated by this query” [Pede00]. The functionality would be reached if all catalogues would have a Z39.50 gateway (server) so that it would be possible to search these catalogues from the same terminal (client). The protocol enables “simultaneously searching of multiple databases” [Pede00]. Both the client and server communicate using Z39.50 protocol messages like depicted in the figure 6 below:

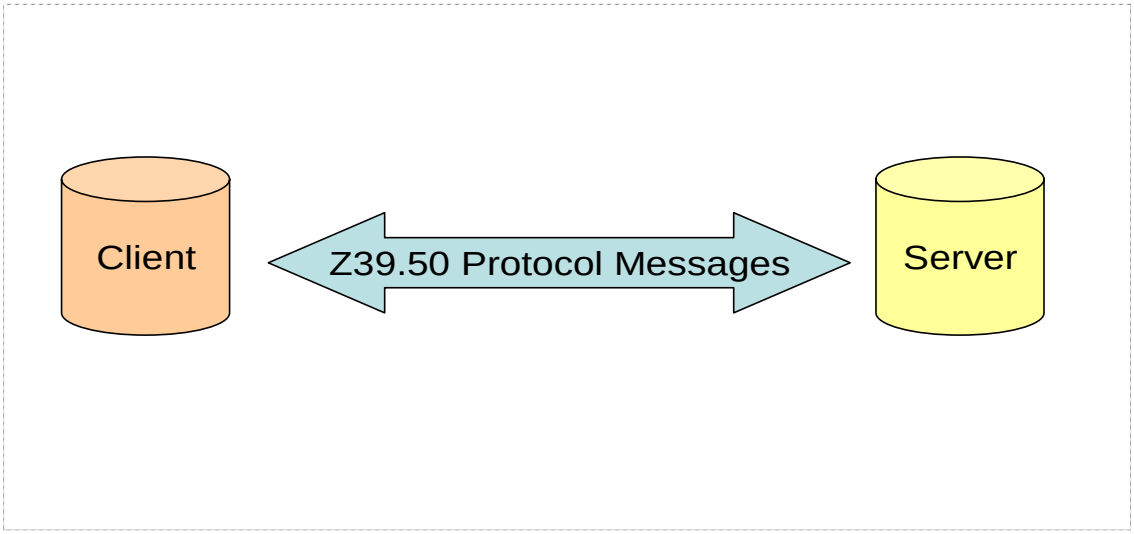

Figure 6 “Z39.50 communication”. Adapted from [Pede00].

Collaboration and resource sharing between libraries can be effectively supported by the use of Z39.50 and at the moment many of the business standard systems for digital libraries support Z39.50 [Pede00]. Z39.50 is designed to facilitate communication between computer systems, mostly like such systems that are used to manage library catalogues [Pede00]. It could be used for communication between the computer of a cataloguer and the library

catalogue itself. The communication could as well exist between a user in the web and the digital library management [Pede00]. This protocol is often used for collaboration between digital libraries. For example, it is possible “to establish a virtual union catalogue of independent and distributed Z39.50 servers” [Pede00].

The difference between Z39.50 and OAI, according to Pederson [Pede00], is that OAI's uses a metadata-harvesting approach, where as the protocol Z39.50 is a search-and-retrieval protocol. Both Z39.50 and OAI provide "federated searching" [Pede00], which enables users “to gather information from multiple related resources through a single interface” [Pede00]. That way one search query can present search results from many resources [Pede00]. This mechanism raises the completeness of the available information so that the user does not have to search several resources separately.

Z39.50 uses an “online, real-time connection between the searcher's system and one or more targets using a thick and complex set of communications protocols” [Pede00]. A query is transmitted to many Z39.50 targets within sessions, and the results from each query are collected, analysed and presented to the user. “The advantages of Z39.50 are its ability to search remote resources through a common user interface and the immediacy with which it accesses current information in real time” [Pede00]. The mechanism of establishing online connections with remote servers makes Z39.50 a subject to the instability [Pede00]. This is a disadvantage of Z39.50, whereas the OAI does not need to establish and maintain such interactive connections among all the original repositories. The OAI creates “pre-built metadata collections” and collects all the metadata it aggregates from the data providers in advance [Pede00].

5.4 The Digital Object Identifier system

The Digital Object Identifier (DOI®) is a product of the International DOI Foundation, a non-profit organization that was founded for the management of the development of the DOI system and for the regulation of the policy and licensing matters of the DOI system [Pask01]. The Digital Object Identifier (DOI®) is “a system for persistent identification and interoperable exchange of intellectual property on digital networks” [Pask01]. The system offers an extensible framework as a solution for intellectual content management. The form and the digital environment of the data do not matter in this case. “By integrating an identifier into a DOI, the identifier becomes actionable as a standard hyperlink (but unlike URL, persistent), and can function in DOI applications across a variety of platforms” [Pask01].

The DOI consists of two components: the prefix and the suffix. Here is an example of a DOI:

10.1000/123456

In this example, the prefix is "10.1000" and the suffix is "123456". The prefix itself has two components a “10” (all DOIs start with "10") and a number (string) that is assigned to the participant that wishes to register the DOI. Separated by a forward slash from the prefix comes a unique suffix (unique to a given prefix). The suffix identifies the entity. The DOI suffix can be any alphanumeric string that the applicant chooses [Pask01].

After its building, the DOI is assigned to a resource [Pask01]. The length of a DOI is not limited. “A DOI may be assigned to any item of intellectual property, which must be precisely defined by means of structured metadata” [Pask01]. Once assigned, the DOI stays persistent and unchanged [Pask01].

The DOI system unites resolution, metadata and policy systems [Pask01]. The resolution system resolves the DOI to a value such as a URL, where as the users do not have to be aware of changes to URLs in order to use DOI [Pask01]. The DOI metadata system provides consistence of mappings

between application areas [Pask01]. DOI policy system provides rules and mechanisms for implementations “by means of a number of Registration Agencies” [Pask01].

Actually, there are two common approaches for naming objects on the Internet: the Uniform Resource Name (URN) and the Uniform Resource Identifier (URI), and DOI conforms to the functional requirements both of them. The difference is that URI and URN specifications deal only with syntax and partly associated implementation through resolution, not with description or persistence policy [Pask01]. Remarkably, general practical implementations of these specifications as object naming do not exist. Both URI and URN are specifications, and not functioning implementations [Pask01]. The DOI is actually a practical implementation of URI and URN [Pask01].

5.5 METS

“The METS schema provides a flexible mechanism for encoding descriptive, administrative, and structural metadata for a digital library object, and for expressing the complex links between these various forms of metadata“ [U.A.04b].

A METS document consists of five major sections by types of the metadata they contain: Descriptive Metadata, Administrative Metadata, File Groups, Structural Map, Behaviour [U.A.04b].

At the beginning of the document comes the METS Header, which allows the user to record minimal descriptive metadata within the METS document about the METS object itself. This metadata includes such information as the date of creation, the date of its last modification, and a status for the METS document. The data about the agents who play some role in regard to the METS document can also be included [U.A.04b]. Here is a small example of a METS Header:

```
    <metsHdr CREATEDATE="2004-11-30T12:30:00"
RECORDSTATUS="Complete">
          <agent ROLE="CREATOR" TYPE="INDIVIDUAL">
            <name>Maria Shustitskiy</name>
          </agent>
    </metsHdr>
```

Following the Header is the first section of a METS document - the Descriptive Metadata section. It consists of one or more <dmdSec> (Descriptive Metadata Section) elements. Each <dmdSec> element may include a pointer to external metadata (an <mdRef> element that provides a URI which may be used in retrieving the external metadata), internally embedded metadata (within an <mdWrap> element, which provides a wrapper around metadata embedded within a METS document), or both [U.A.04b].

Administrative Metadata <amdSec> elements contain the administrative metadata related to the files including a digital library object, as well as metadata related to the original source information that was used to create the object [U.A.04b]. The administrative metadata has four main forms in a METS document: Technical Metadata (information concerning the creation of a file, format, and characteristics of its use), Intellectual Property Rights Metadata (information on copyright and licensing), Source Metadata (descriptive and administrative metadata about the analog source of origin of a digital library object), and Digital Provenance Metadata (information concerning file relationships) [U.A.04b].

The next section is the File Section (<fileSec>), which contains one or more <fileGrp> elements that are used to assemble related files. A <fileGrp> element provides a list all of the files which contain a single digital version of the digital library object [U.A.04b].

The fourth section of a METS document is the Structural Map Section [U.A.04b]. It serves the navigation of users in the digital library object and defines its hierarchical structure. “The <structMap> element encodes this

hierarchy as a nested series of elements” [U.A.04b]. Each includes attribute information stating the information about the division. A may also contain several METS pointer (<mptr>) and file pointer (<fptr>) elements for identification of the content corresponding with that division.

METS pointers indicate separate METS documents that contain the relevant file information for the division. This is helpful when large collections of material are encoded, so that the size of each METS file in the set is kept small. File pointers identify files (or groups of files) that correspond to the segment in the hierarchy represented by the current within the <fileSec> section of the METS document [U.A.04b]. Related to the Structural Map section is the Structural Links section of the METS format. This section contains only a single element, <smLink>, which can be repeated. This section allows recording the existence of hyperlinks between items within the structural map. This is useful when applying METS to archive web sites, and maintaining a documentation of the hypertext structure of the sites outside of the HTML files [U.A.04b].

The last section of a METS document is the behavior section. This section is for associating executable behaviors with content in the METS object [U.A.04b]. This section comprises one or more <behavior> elements. The elements have an “interface definition element that represents an abstract definition of the set of behaviors represented by a particular behavior section” [U.A.04b].

The METS schema offers a constructive standard for the exchange of digital library objects between repositories. In addition, METS provides the ability to relate a digital object to actions or services [U.A.04b]. METS attempts to provide an XML document format for “encoding metadata needed for management of digital library objects within a repository and exchange of such objects between repositories (or between repositories and their users)“ [U.A.04b].

5.6 OpenUrl

The OpenURL is a linkage system that is created to facilitate the transfer of the metadata from the information source to a service component. The task of the latter is to provide “context-sensitive services for the transferred metadata” [SHBA00].

Reference linking is a service built on the metadata. OpenURL enables linking from an information source to the service component. This linking is called the outbound linking [SHBA00]. OpenUrl provides the delivery of context-sensitive services for metadata. The providers of information have to add an OpenUrl to the metadata shown as an outcome of a search/browse action [SHBA00].

An OpenURL permits the delivery of metadata from an information resource to a service component. An OpenURL consists of two basic elements: the Web address of the relevant service component the base called URL and the content [Sade01]. The content is for example the metadata of the particular object; or it could be reference to the metadata object [Sade01]. The provider of the context-sensitive services can be any third party, and the way in which they are provided is through a service component, or "link server” [Sade01].

The figure 7 below shows the general OpenUrl process flow.

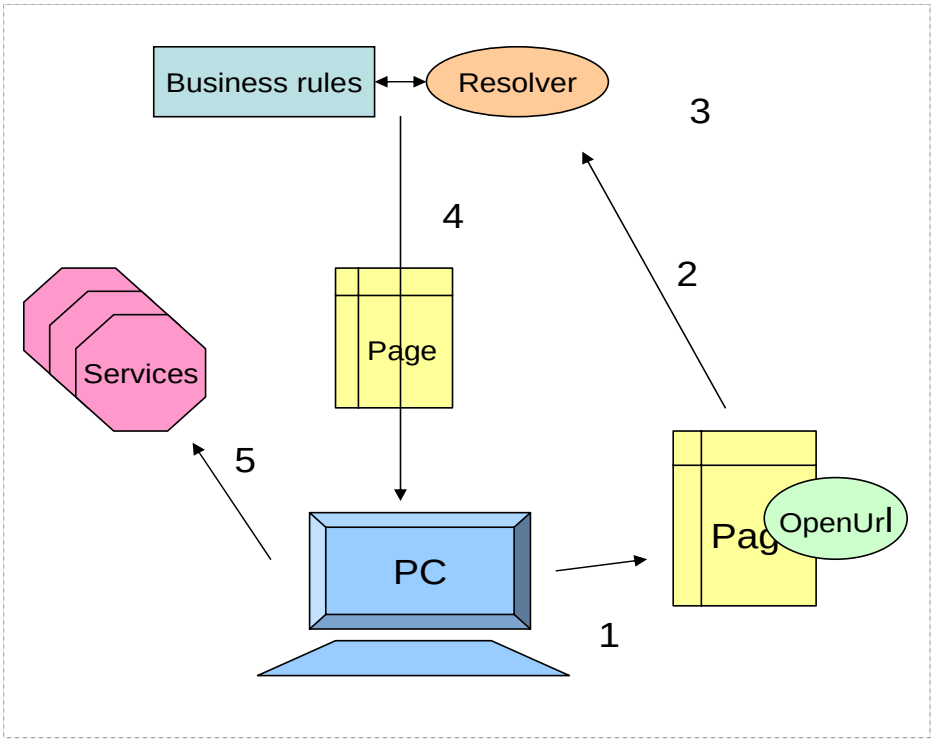

Figure 7 “OpenUrl Process Flow”. Adapted from [U.A.04i].

The OpenUrl Process Flow includes the following steps:

First, a user or user agent, called requestor, accesses an html page that contains an OpenURL [U.A.04i]. The OpenURL was placed by a third party service providers that want to allow the requester to choose to receive service options related to the corresponding object [U.A.04i].

Second, the requestor activates transfer of the encoded metadata in the OpenURL to the service component = resolver specified by the OpenURL [U.A.04i]. An OpenURL is activated by clicking a link. The encoded metadata is transmitted to the resolver [U.A.04i].

Step three, the resolver interprets the encoded metadata based on business rules maintained in its local information service environment and identifies one or several services to be returned to the requester [U.A.04i].

Step four, the services are returned to the requester through an html page [U.A.04i]. And step five, the requester selects one or more services by clicking on the provided links [U.A.04i].

OpenUrl is important for digital library collaboration because it enables context-sensitive services for the transferred metadata, which leads to a higher effectiveness of the collaborative work of digital libraries as well as to more value-added services for the users.

The OpenURL framework can support a range of different service components. One of them is SFX Server. SFX Server is the “Ex Libris manifestation of a service component” [Sade01] and provides the service of a context-sensitive reference linking [Sade01]. The input must have the form of an OpenURL consisting of metadata objects and identifiers. The metadata is then taken from the OpenURL and demanded from the original resource or from another resource by means of the identifier [Sade01]. After this, the SFX Server analyzes it and “generates a set of extended services appropriate to the specific metadata and the specific user” [Sade01]. “With SFX, libraries can define rules that allow SFX to dynamically create links that integrate their information resources regardless of who hosts them - the library itself or external information providers” [Sade01].

5.7 RDF

Resource Description Framework (RDF) is a foundation for processing metadata. RDF provides interoperability between computer applications on the Web; it enables automated processing of Web resources. The purpose of creation of RDF is to define a domain neutral method for resources descriptions on the Web, to which a particular application domain and its semantics would be without importance [LaSw99]. The mechanism should at the same time be appropriate for describing information about any domain. RDF offers a schema system to facilitate the definition of metadata. The RDF schemas are written in

RDF [LaSw99]. RDF also supports the “sharability” of schemas and the reusability of metadata definitions [LaSw99].

The base element of the RDF model is the triple [Cham01]. In the figure 8 below you will find such a triple.

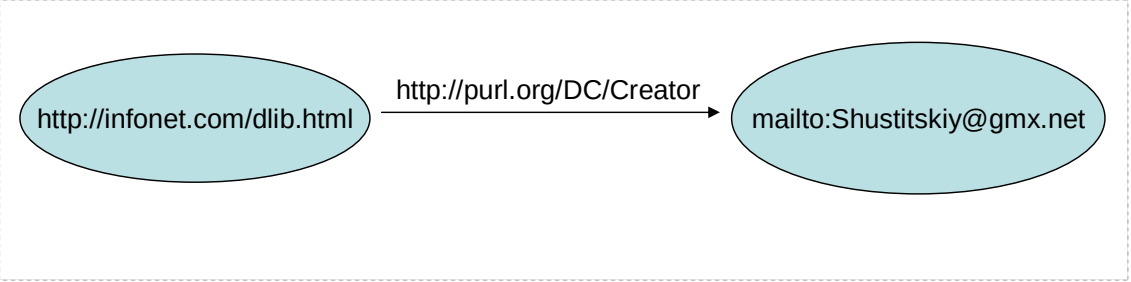

Figure 8: “A triple”. Adapted from [Cham01].

A triple consists of tree parts: a subject, an object and a predicate [Cham01]. All three are resources. The subject is linked to the object through an arc labeled with a predicate [Cham01]. So, the subject has a property (predicate) valued by the object [Cham01]. For example, the triple in figure 8 could be read as ``Shustitkiy is the creator of dlib.html''.

RDF can be used in many ways. It is useful in resources detection by offering better search engine capabilities and in cataloging for a better content description on a Web site. It is also used by “intelligent software agents to facilitate knowledge sharing and exchange” [LaSw99], in “content rating”, in descriptions of groups of pages referring to each other, for descriptions of intellectual property rights [LaSw99]. It is also helpful for indication of the “privacy preferences of a user as well as the privacy policies of a Web site” [LaSw99]. RDF is a very important development for digital libraries, since it offers so many ways, in which it can be used, and all of them are essential for collaboration.

6 Current collaborative projects

The field of digital libraries is an area, where all participants gain from coordinated, collaborative activities. This field offers very interesting opportunities for collaboration as globally distributed resources created by different groups engaged in collaborative efforts are the most content-rich collections [WRGr99]. The existence and the level of the collaboration of digital libraries of a country within the country or among several countries is determined mainly by the criteria of the significance of information for these countries [ZABE01]. The formation of unified library information web is important not only due to the significance of economy of financial and technical means, but also due to the necessity of information exchange within a country and beyond its boundaries [ZABE01].

Unfortunately, there are uncoordinated efforts in many countries, while cooperative programs of research and intellectual infrastructure development can help evade replication of efforts, prevent the development of disorganized digital networks, and encourage interchange of scientific knowledge on an international basis [WRGr99].

In this section we shall provide some examples of significant projects on library collaboration.

6.1 The Joint NSF/JISC International Digital Libraries Initiative

The early digital library projects through the 1990s were largely experimental activities [PaWr02]. The research in this area was often sponsored by the U.S. National Science Foundation (NSF) and the U.K. Joint Information Systems

Committee (JISC) [PaWr02]. In this period, many important advances in digital library techniques were achieved. In 1999 an important step for the development of digital libraries and their collaboration was made: a three-year International Digital Library Initiative emerged from linking of the digital library research program of NSF with the similar program of JISC [PaWr02]. The aims of International Digital Library Initiative were to:

- “identify a collection of information which is not accessible or usable because of technical barriers, distance, size, system fragmentation or other limits
- using this as a test-bed, create the understanding and new technology to make it possible for such information to be found, delivered to, and/or exploited by, a distributed set of users; and
- evaluate the effect of this new technology and its international benefits”. [WRGr99].

The initiators saw their goal in “enabling users to access digital collections more easily and supporting broader use of these collections” [WRGr99]. As a result, six projects on digital library research were recommended for funding, and the participating parties have been extremely pleased with the outcome of the joint work and the quality of the projects, which were supported. The six projects were:

- “Cross-Domain Resource Discovery: Integrated Discovery and use of Textual, Numeric and Spatial Data” University of California, Berkeley / University of Liverpool
- “HARMONY: Metadata for resource discovery of multimedia digital objects” Cornell University / ILRT / DSTC
- “Integrating and Navigating ePrint Archives through Citation-Linking” Cornell University / Southampton University / Los Alamos National Laboratory
- ‘Online Music Recognition and Searching (OMRAS)” University of Massachusetts / King's College, London

- “Emulation options for digital preservation: technology emulation as a method for long-term access and preservation of digital resources” University of Michigan / CURL
- “The IMesh Toolkit: An architecture and toolkit for distributed subject gateways” University of Wisconsin-Madison / UKOLN / ILRT

These six research projects were funded during the first phase of the Digital Library Initiative [U.A.04c].

The intention of the Digital Libraries Initiative Phase II is to extend the research undertaken during the first phase. The second phase is sponsored by NSF; the Library of Congress joins the Defense Advanced Research Projects Agency (DARPA), the National Library of Medicine (NLM), The National Aeronautics and Space Administration (NASA), and the National Endowment for the Humanities [U.A.04c]. During Digital Libraries Initiative - Phase II a large number of projects and workshops on digital library development has been sponsored [U.A.04c]. Some of them are:

- “High-Performance Digital Library Classification Systems: From Information Retrieval to Knowledge Management “ by University of Arizona
- “Re-inventing Scholarly Information Dissemination and Use” by University of California Berkeley
- “A Multimedia Digital Library of Folk Literature” by University of California Davis
- and others.

6.2 The University of Washington Digital Libraries Initiative

The University of Washington has founded a Digital Libraries Initiative, which provides a framework for “effective and broad collaboration among faculty, engineers, students and librarians” [BuZi99]. The aim of the initiative was to offer a center point for the “creation, use, and investigation of electronic information services, resources, and systems” [BuZi99]. The organization of the digital collections and resources was developed from a user-oriented perspective. For the University of Washington Digital Library Initiative the collaboration with its partners is of fundamental importance to their work. The partners of the University of Washington in the Digital Library Initiative were the University Libraries, the departments of Electrical Engineering and Technical Communications in the College of Engineering, and the School of Library and Information Science. All of the partners have unique and complementary roles. As a result of a cooperative effort between the libraries and the Center for Information Systems Optimization was the development of CONTENT, the high performance image archiver [BuZi99]. “Faculty in the School of Library and Information Science are interested in partnering in digital library development and contributing to the users’ understanding of retrieval issues” [BuZi99]. They are interested in the use of cognitive sciences in understanding the learning processes of both non-specialist and expert searchers. Continuous improvement of the systems requires the undertaking of a rigorous research on the information-seeking behaviors of the largest user-group, the undergraduates. The “best to date” program of the Libraries collaboration at the University of Washington is the UWired Program, a joint program of the Libraries, Computing & Communication, the Office of Undergraduate Education and, more recently, Educational Outreach. “UWired builds student and faculty skills with technology, and serves as a model program for focusing resources in a collaborative setting to support individual and interdisciplinary technology needs” [BuZi99].

Another project of the Digital Libraries Initiative of the University of Washington is the University of Washington Information Gateway. The Gateway provides a focal point for the migration of all electronic resources, whether commercially acquired or locally developed [BuZi99]. The aim is to provide a broad and useful access to them. For this, the Libraries have built an “SQL-based digital registry to which metadata is mapped from the Innovative Interfaces library catalog” [BuZi99]. Another topic of the collaboration is the providing of Web access to commercial digital publishings, abstracting/indexing databases, catalogs and websites by the Digital Libraries initiative in common with other universities. To extend this, the Libraries have begun to digitize the unique holdings and primary sources held by the Libraries and by various university departments. The collections are being displayed and managed with the image archive package known as CONTENT [BuZi99].

6.3 The British Library Digital Libraries program

The British Library has a long tradition of collaboration and partnership since its establishment in 1974. There are well-established collaborative services and schemes in all its directorates, there is also a wide range of new initiatives under development with partners. The main collaborative efforts take place in the field of the intake and record creation for legal deposit materials with the other five legal deposit libraries. The British Library is also involved in national and international projects such as NEWSPLAN, ESTC, ISTC, The European Library and Gabriel and it is working with international organizations such as CERL. Close collaboration and partnership with other libraries takes place mostly in such areas as “collection development, preservation and retention, access, bibliographic services and record creation” [U.A.03a]. As a result of strong emphasis on collaboration, and to strengthen its commitment to working in partnership with others in the library and information sector, the Library established a Co-operation and Partnership Programme, whose role is to “define and develop the programme of collaboration and partnership in the key areas mentioned, to coordinate existing and new British Library collaborative

activities and to administer a newly created fund to support collaborative and partnership initiatives” [U.A.03a].

6.4 The European Library Project

"The European Library" (TEL) was an initiative in a form of a consortium of national libraries, which has been founded with the aim of improving collaborative work of Europe's national libraries in the area of digital and other collections [U.A.03b]. The project officially started on 1 February 2001 and was effectively running for three years. The objective of TEL was to set up a cooperative framework with a system for access to the “major mainly digital collections in European national libraries” [U.A.03b].

The partners in the consortium were The British Library (Co-ordinating Partner), Die Deutsche Bibliothek (Germany), Koninklijke Bibliotheek (Netherlands), Helsinki University Library (Finland), the Swiss National Library, Biblioteca Nacional (Portugal), Biblioteca Nazionale Centrale Firenze (Italy), Istituto Centrale per il Catalogo Unico (Italy), Narodna in Universzitetna Knjiznica v Ljubljani (Slovenia) and the Conference of European National Librarians (CENL) [U.A.03b].

TEL was expected to develop collaborative approaches for access to “large-scale content in technical and business issues” [U.A.03b] and to contribute so to the cultural and scientific knowledge infrastructure within Europe.

Following results were to be achieved through four specific objectives:

- “The establishment of cooperative approaches to business, licensing, copyright matters and public relations.
- A second objective is the development of business plans and models for the maximization of the benefits of collaboration.

- The third objective is the metadata development. A concerted best practice approach to metadata standards and schemas has to be found in collaboration with relevant projects and agencies. The result of this objective would be an agreed protocol between the members for metadata construction, reduction in the number of schemas in use overall, and schemas for agreed approaches to be tested in the testbed work package.

- The forth objective was the development of interoperability testbeds with an accent on the collaborative aspects. The testbeds will be the environment to jointly experiment and provide optimal solutions for the development of services. Two testbeds were expected to be developed focused on Z39.50 and XML" [U.A.03b].

After a three-year work, the main achievements of the TEL project were:

- "the creation of a consensus between national libraries on the mission and content of The European Library and agreement on a business model for the development and support of The European Library Service,

- the establishment of agreed metadata profiles (based on common standards) for objects and collections within an environment, which will enable their controlled evolution to accommodate new types of material and new collections as The European Library grows and

- "the design of a flexible system architecture based on the SRU protocol" [U.A.03b].

6.5 Digital Cultural Heritage Community of East Central Illinois

The Digital Cultural Heritage Community project (DCHC) was founded by the University of Illinois Library at Urbana-Champaign with the aim to create an online database that would comprise digitized materials from museums, libraries archives and schools for use by elementary school teachers in their classrooms [U.A.04d]. It was developed and worked out in a face-to-face and online collaboration between diverse institutions in the East Central Illinois. The overall goal of the project was to build and evaluate a model of collaborative digital environment that would provide elementary classrooms with access to multimedia information on topics that addressed Illinois State Board of Education Learning Standards for Social Science. The project was administered by the Digital Imaging and Media Technology Initiative at the University of Illinois Library at Urbana-Champaign. The museum partners of the project included the Early American Museum and the McLean County Museum of History, as well as the Illinois Heritage Association, and the library partners included the Lincoln Trail Libraries System and the Rare Book and Special Collections Library of the University of Illinois [U.A.04d]. As the project progressed, other museums and libraries took part in the project at different levels, either by attending workshops or adding data to the online database. Three elementary schools also participated, with concentration on third, fourth and fifth grade classrooms. [U.A.04d].

The DCHC project was based on the concept of a digital community, where institutions would contribute to a database images, text, other multimedia objects and descriptive information addressing common themes. Teachers would then use the database to engage their students through more robust lesson plans. The project aimed at making it easy for elementary school teachers to utilize given resources, enabling incorporation of online materials into their classroom activities in meaningful ways for their students. The online database was set up using the Dublin Core (DC) metadata schema [U.A.04d].

6.6 Collaborative projects of the U.S. Department of Education

U.S. Department of Education was created in 1980 by bringing together offices from several other U.S. departments. Its mission is to “ensure equal access to education and to promote educational excellence throughout the USA” [U.A.04e]. At the moment the Department is supporting several projects in the field of inter-institutional collaboration and exchange of digital information. For example, the Department of Education is funding a collaborative project involving the University of Chicago, Columbia University, and the Triangle South Asia Consortium in North Carolina to produce and distribute electronic dictionaries for each of the twenty-six modern literary languages of South Asia. This project is expected to enlarge the international lexical infrastructure and supply high quality resources for use in education and by readers [U.A.04e].

The Department of Education is also funding a collaborative project, which involves the Center for Research Libraries, Chicago, and Columbia. The project aims at preparing and bringing about a broad array of digital resources on South Asia [U.A.04e]. The Digital South Asia Library offers digital materials for reference and research on South Asia to numerous users. Other libraries committed to join in the project are the British Library, the University of Oxford's Indian Institute Library, Cambridge University Library, the Roja Muthiah Research Library (Madras), and the Urdu Research Centre (Hyderabad) [U.A.04e].

6.7 Chinese Memory Net

Chinese Memory Net (CMNet) is a collaborative research project directed at the creation of a global digital library in Chinese studies. The project started in 2000 and ran for 3 years. It is a result of collaborative work of the Simmons College, Tsinghua University, Peking University in Beijing, Jiao-Tong University in Shanghai, National Taiwan University, National Tsinghua University and the Academia Sinica. The project was expected to bring selective academic educational and research partners in the United States, Taiwan, and China together working toward an effective and sustainable global digital library in Chinese Studies [U.A.04g].

Each participating institution possesses valuable collections of Chinese culture and heritage. These collections are unique and essential to education and research, but most of them are currently not accessible or usable because of the remoteness, their form or technical obstacles. This research project is aimed at finding new ways to allow academic users to access and use these significant research collections through global networks [U.A.04g].

The direct objectives of the project are “to prevent the development of fragmented and non-interoperable digital systems in Chinese studies already taking place in different parts of the world and so to avoid duplication of efforts” [U.A.04g]. It is also important for the project “to define characteristics of an interoperable system in dealing with Chinese information resources, to find possible solutions for multilingual (specifically Chinese) language processing, and to develop commonly accepted standards of methods, protocols, tools and technology for system development in the current networked environment” [U.A.04g].

This research project is also aimed at exploring the use of current networks of different speeds, bandwidth, and configurations for digital data and image transmission, which are available to project partners. The result should be the enhanced education and research in Chinese studies, commonly agreed standards and tools, and the information achieved from this research [U.A.04g].

6.8 African Online Digital Library

The Multi-Lingual Digital Library for West African Sources or the African Online DigitalLibrary (AODL) was created by the Michigan State University together with NSF - Institut Fondemental d'Afrique Noire (IFAN) and West African Research Center (WARC) (US-Africa). MATRIX, working in cooperation with the African Studies Center at MSU, and in partnership with premiere research institutions in Africa, is pioneering the African Online Digital Library [U.A.03c]. The goal of this online digital repository is “to adopt the emerging best practices of the American digital library community and apply them in an African context “ [U.A.03c].

AODL offers many benefits for a wide variety of scholars and institutions by producing multilingual, multimedia materials for both scholarly research and public viewing audiences [U.A.03c]. AODL serves scholars and students conducting research and teaching about West and South Africa as well as teachers and students of African languages. It also provides a valuable model for creating and distributing a wide spread of materials in a region with very limited electronic possibilities [U.A.03c].

This project’s aim is to build a multi-media digital library of West African sources in multiple languages. Collaborating organizations include several organizations in Michigan, USA, the Institut Fondemental d'Afrique Noire (IFAN) and West African Research Center (WARC) in Dakar, Senegal [U.A.03c]. Four types of material will be archived including historical manuscripts from the 19th and early 20th centuries and other materials. The project will attempt to identify and provide metadata that describe the origin and content of this digital information. A multiple-languages interface will allow users to access to content through the use of standardized subject headings in the original languages. “Caching proxy server technology will be used a local site servers in Senegal to address low-speed international network connections common in African countries”

[U.A.03c]. The project is expected to serve as a “model for larger-scale efforts by African and US researchers to create new repositories for the broader research and education communities” [U.A.03c].

7 “Open Access” projects and initiatives

“Open Access” is a new model of information exchange on the Internet, that gains more and more importance nowadays. There are several initiatives working on it at the moment, and there already exist numerous projects, using the model for the information dissemination. Open Access is also used in many collaborative research projects of digital libraries. Here, the content produced is made openly available, so that other digital libraries can make use of it and consider it as a basis for further research.

7.1 Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities

The Berlin declaration 2003 is a document composed and signed initially by over 30 German and international research organizations and institutes (since then, the number of signatures has grown). It proposes a model of knowledge dissemination, where information is made “widely and readily available to society” [BGEi04]. The undersigned regard themselves as scientists, whose goal is knowledge dissemination. They define internet as an instrument for information distribution, and see open access as the only possible model of fair and global knowledge representation [BGEi04].

It is being underlined, that a perfect functionality of the open access model would require an active participation of all scientists or anyone producing scientific knowledge. The generally available information could include “original scientific research results, raw data and metadata, source materials, digital

representations of pictorial and graphical materials and scholarly multimedia material" [BGEi04].

The authors see two conditions for the open access contributions:

- The author/right holder grants a full and free access to the published information. The reader is allowed to use the information in any way, as long as it does not serve his commercial goals [BGEi04].
- A full version of the work and related materials is being stored in at least one online repository. The repository has to use technical standards that are appropriate for the possibility of open access and has to be established and supported by an institution that is willing to enable open access to all of the users [BGEi04].

In conclusion, the publishers of the declaration state that they also see their goal in promotion of the open access model for the information exchange. They intend to do it by encouraging the scientists and publishers to provide the results of their work for open access, by maintaining the information quality standards and by supporting open access through technical means [BGEi04].

7.2 Budapest Open Access Initiative

The Budapest Open Access Initiative is a worldwide initiative, which arose from a meeting, which took place in Budapest on December 1-2, 2001. The meeting was convened by the Open Society Institute (OSI) [CCEF02]. OSI expected that the meeting would contribute to the free availability of the research articles on the Internet. The meeting resulted in the Budapest Open Access Initiative (BOAI), which is "at once a statement of principle, a statement of strategy, and a statement of commitment" [CCEF02]. The aim of the Initiative is the distribution of peer-reviewed articled through Internet and the open access of any users to this literature. Open access means the "free availability" [CCEF02] of the literature in the internet. The users are allowed to "read, download, copy,

distribute, print, search, link to the full texts of these articles …or use them for any other lawful purpose, without financial, legal, or technical barriers" [CCEF02]. The initiators see the "authors control over the integrity of their work and the right to be properly acknowledged and cited" [CCEF02] as the only restriction, which should be seriously paid attention to.

The promoters of the Initiative expect that the initiative would help to "accelerate research, enrich education, share the learning of the rich with the poor and the poor with the rich… and lay the foundation for uniting humanity in a common intellectual conversation and quest for knowledge" [CCEF02].

The Budapest Open Access Initiative was signed by the participants of the meeting in December 2001. It is still available for signing, and until now it has been signed by numerous scientists, research institutes and individuals, universities, foundations, journals etc. as well as by similar open access initiatives [CCEF02].

The initiators recommend two complementary strategies, which could be used in order to make open access to scientific literature possible:

1. Self-Archiving: authors can put their articles in "open electronic archives" [CCEF02] – e.g. publicly accessible websites. Separate archives will be seen as one by the search engines and similar tools if the technical standards of these archives match those of the Open Archives Initiative ("a protocol for collecting metadata about data files residing in separate archives") [CCEF02]. In this model, the scientists hold the copyright to themselves [CCEF02].
2. Open-access Journals: this is the second recommended strategy, where scholars might create open access journals and put articles there. The copyright will either be retained by authors or it will be transferred to the open access journals [CCEF02]. In any case open access to the documents will be provided.

The objective of the BOAI is to enable a free and open access for all readers to the peer-reviewed literature that the authors wish to give away for free. Their goal is to remove access barriers to such writings and make literature as useful as it can be [CCEF02].

Practically BOAI contributed to the promotion of the open access model by creating the Directory of Open Access Journals [CCEF02], which provides an open access to scientific and scholar journals.

7.3 Directory of Open Access Journals

Directory of Open Access Journals is a service established and funded by the Budapest Open Archives Initiative. It enables access to “quality controlled Open Access Journals” [Jorg04]. The Directory includes scientific and scholarly open access journals, which hold full text articles and use a quality control system able to guarantee the provided content [Jorg04]. A journal can only be included in the Directory of Open Access Journals if it provides quality control on the papers it submits “through an editor, editorial board and/or a peer-review system” [Jorg04].

The implementation process of DOAJ was split in two phases. The first phase was the implementation of Directory itself. It began at the Budapest meeting in December 2001, which resulted in the Budapest Open Access Initiative [Jorg04]. The second phase began in autumn 2003. In this phase the development of a “search system for article-level content” based on metadata [Jorg04] took place.

The Directory’s major aim is an increase of the “visibility and ease of use” [Jorg04] of open access journals.

7.4 The Public Library of Science

“The Public Library of Science … is a non-profit organization of scientists and physicians committed to making the world's scientific and medical literature a freely available public resource” [Sieg04]. The Public Library of Science aims at providing unrestricted and free access for any interested users to the scientific literature of the world. By doing so, it intends to facilitate research and education and to move forward the scientific progress [Sieg04].

The Public Library of Science implements a new business model, essential for realizing the potential of the open access. The model treats the costs of publication as “the final integral step of the funding of a research project” [Sieg04]. The Library publishes its own open access journals to prove that this model is effective. The journals are run by professionals and publish peer reviewed papers in the fields of medicine and biology [Sieg04].

Unlike the Budapest Open Access Initiative, the Public Library of Science offers its services not in all academic fields, but only in the area of science [Sieg04]. Still the two initiatives actively support each other.

The Library states its core principles as follows: open access, excellence, scientific integrity, breadth of scope, cooperation with similar organizations, financial fairness, community engagement, internationalism, and science as a public resource [Sieg04].

The Library is being funded by grants from several foundations and institutes, as well as by donations from universities, private citizens etc. [Sieg04].

7.5 Create Change

“Create Change” is an initiative concerning the transformations in the system of the worldwide information dissemination [CDJo04]. The initiators address the problem of the scholarly communication, which arose due to the continuously increasing volume and cost of the scientific and scholar literature. Following the authors, the problem is that libraries cannot afford to offer all of the world scientific printings to their users and have to restrict their holdings to a small fraction of the world supply. As a result, users have only access to a very limited number of printed articles, and authors of the articles can only communicate to a very restricted number of readers [CDJo04].

The initiators call this problem a crisis in scholarly communication and suggest a development of a new system, which can meet the needs of the authors and the readers [CDJo04].

The goal is “to make scholarly research as accessible as possible to scholars all over the world, to their students, and to others who might derive value from it” [CDJo04]. In order to reach this goal, the initiators suggest that scholars take back the control of scholarly publication from commercial publishers, to persuade the publishers to alter their primary goals to “the widest possible dissemination of scholarly information”, and to create alternatives to commercial scholarly publications [CDJo04].

8 Conclusion and the future directions in digital library collaboration

Technology has fundamentally changed the world of publishing and the rules according to which authors, publishers, libraries, and users interact with each other. Digital libraries are strongly involved in this development. During the phase of transition from the world of print information materials to digital information there is a possibility to analyze shifting costs, to design new economic models and to adjust legislation [NWDH96].

The future of digital libraries cannot be seen without extensive collaboration among them. Collaboration is the main chance of the digital libraries to move forward, to increase their efficiency, to provide more services and information for the users and to develop technologically. Collaboration will be beneficial for all libraries as all collaborators possess their own experience, knowledge, expertise and capabilities. Ways to establish collaboration depend on many factors and will have to be found in all cases individually.

At the moment, the field of digital library collaboration is a large area for extensive research activities. As Lynch and Garcia-Molina [LyGM95] noted, digital libraries are not simply technological constructs; they exist within a rich legal, social, and economic context, to which they have to be adopted continuously. Digital libraries have to pay attention to these restrictions and meet the needs of this background. In view of these requirements we will need research on the topics of rights management, economic models for the use of electronic information and support for those economic models [LyGM95]. It is important to remember that digital libraries and collaboration models have to be adapted to the economic background and the structure from the point of view of the funds allocation, as well as to the social collaborative situation [FMLM03]. The legal issues of digital library collaboration comprise the matters of the material sharing, the control over the ownership, the privacy management, the material and reuse etc [FMLM03]. The challenges become more complex when

collections become multi-national [FMLM03]. These matters can most efficiently be handled with the support government and related organizations, as joint agreements on legal issues have to be made [FMLM03].

Effective transformation of structures, models and mechanisms and achievement of economic and social prosperity will only be possible in case of deep understanding of the driving forces of change and by addressing problems in an organized, practical and global way [NWDH96].

The future of digital libraries is marked by significant uncertainty, a lot of which can be classified as issues of intellectual property and economics. Many open questions exist, both about what is technologically possible, and what will actually happen [NWDH96]. The results of scientific research and development work are already being widely applied and implemented.

There are several key trends that can be recognized today and for the nearest future:

Digital library collaboration will probably underlie mostly unitary standardization or widely accepted best practices. In the next years the main task of organizations responsible for digital libraries will be to find successful approaches for such topics as metadata encoding, rights management, preservation, ubiquity and other processes necessary for the field of digital library collaboration. When those techniques and approaches advance and find popularity, the focus will switch from them to other research issues [PaWr02].

The main existing standard organizations will have to agree on common techniques and approaches and software producers should apply these agreed-upon standards in their products. After that, the focus of digital library development is expected to shift from experiments and trials to implementation of achieved results [Pede00].

In the next stage of digital library collaboration development the topic of providing greater efficiency for library users, increased interoperability among digital libraries, and more alternatives for individual preferences will gain importance [PaWr02].

The most revolutionary trend is the movement towards the “open access” model of digital library collaboration and information exchange. This model proposes the scientists and authors of articles to enable a free and unrestricted access to the printed materials and to treat costs as a final step in the research projects [CCEF02]. Of course, we don’t have to expect that all the materials will be made accessible for everyone with no constraints and no charge. But this is a model, which will suit the aims of many scientists and publishers, and which will probably become very popular.

9 References

[AnVi99] Antopolskiy A.B., Vigurskiy K.V. „The concept of electronic libraries“, Russian Digital Libraries Journal - 1999 - Vol. 2 - No 2. http://www.iis.ru/el-bib/1999/199902/antopol/antopol.ru.html

Requested on January 28, 2004.

[Arms99] Arms, William, Y: Digital libraries”, 1999. http://www.cs.cornell.edu/wya/DigLib/MS1999

[Arms01] Arms, William, Y.: “Digital libraries”, 2001.

[BBCC03] Birmingham, Bill; Böhm, Klemens; Christophides, Vassilis; Cluet, Sophie: “EU-NSF Digital Library Working Group on Interoperability between Digital Libraries”. http://www.iei.pi.cnr.it/DELOS/NSF/interop.htm Requested on January 31, 2004.

[BGEi04] Bullinger, Hans-Jorg; Gaethgens, Peter ; Einhäupl: “Berliner Erklärung über offenen Zugang zu wissenschaftlichem Wissen“, Oktober 2, 2003. http://www.mpg.de/pdf/openaccess/BerlinDeclaration_dt.pdf Requested on June 5, 2004.

[Bree02] Breeding Marshall: “The emergence of the Open Archives Initiative”. http://www.openarchives.org. April, 2002. Requested on January 4, 2004.

[BSPi02] Bennett Nuala, Sandore Beth, Pianfetti Evengeline: “Illinois Digital Cultural Heritage Community - Collaborative Interactions Among Libraries, Museums and Elementary Schools” in D-Lib Magazine, January 2002. http://www.dlib.org/dlib/january02/bennett/01bennett.html

Requested on September 4, 2003.

[BuCo03] Butler, Timothy; Coleman, David: “Models of Collaboration”, September 2003. http://www.collaborate.com/publication/newsletter/publications_newsletter_september03.html

[BuZi99] Bunker Geri, Zick Greg: „Collaboration as a key to digital library development“ in “D-Lib Magazine” March 1999 http://www.dlib.org/dlib/march99/bunker/03bunker.html Requested on May 23, 2003.

[CDJo04] Case, Mary; Davis, Mary Ellen; Johnson, Rick: “About the Create Change”. http://www.createchange.org Requested on August 8, 2004.

[CCEF02] Chan, Leslie; Cuplinskas, Darius; Eisen, Michael; Friend, Fred;…: “Budapest Open Access Initiative”, February 14, 2002. http://www.soros.org/openaccess/read.shtml Requested on June 5, 2004

[Cham01] Champin, Pierre-Antoine. “RDF Tutorial”, April 5, 2001. http://www710.univ-lyon1.fr/~champin/rdf-tutorial/

[CHEN03] Chen Chingchih: „The Chinese Memory Net“. http://www.dli2.nsf.gov/internationalprojects/sino/CMNet.pdf Requested on August 20, 2004

[Coll94] Collier Mel: “Defining the electronic library” in “International Conference on Electronic Library and Visual Information Research <5, 1994>: ELVIRA 1: papers from the third ELVIRA conference”. May 1994.

[EnFe00] Endres, Albert; Fellner, Dieter W.: „Digitale Bibliotheken“, 2000.

[FMLM03] Fox Edward, Moore Reagan, Larsen Ronald, Myaeng Sung, Kim Sung:” Toward a Global Digital Library” in D-Lib Magazine,

October 2002 http://www.dlib.org/dlib/october02/fox/10fox.html
Requested on September 3, 2003.

[GoZa99] Golenskiy S.P., Zaharov V.P.: “Information and Technologies” Russian Digital Libraries Journal - 1999 - Vol. 2 - No 2. http://www.iis.ru/el-bib/1999/199902/golensky/golensky.ru.html. Requested on January 28, 2004.

[Hack03] Hacken, Richard: “Types of digital library cooperation in German studies”, a paper delivered at the ACRL/WESS program, June 23, 2003. http://www.lib.byu.edu/estu/wess/torontotalk.html Requested on March 15, 2004.

[Herz98] Herzberg, Amir: “Safeguarding digital library content”, January 1998. http://www.dlib.org/dlib/january98/ibm/01herzberg.html

[Hill01] Hillmann Diane: “Using Dublin Core”. April 12, 2001. http://dublincore.org/documents/2001/04/12/usageguide/. Requested on January 12, 2004.

[Jorg04] Jorgensen, Lotte: “Directory of Open Access Journals”. http://www.doaj.org/articles/about

Requested on June 4, 2004.

[Kunz99] Kunze, J.: “Encoding Dublin Core Metadata in HTML” , December 1999. http://www.ietf.org/rfc/rfc2731.txt

[Lago02] Lagoze, Carl: “The Open Archives Initiative Protocol for Metadata Harvesting, Protocol Version 2.0” September 13, 2002. http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm Requested on February 2, 2004.

[LaSw99] Lassila Ora, Swick Ralph R.: **“**Resource Description Framework (RDF) Model and Syntax Specification.” February 22,1999. http://www.w3.org/TR/1999/REC-rdf-syntax-19990222. Requested on January 15, 2004.

[Lein98] Leiner Barry M.: “The Scope of the Digital Library” October 15, 1998. http://www.dlib.org/metrics/public/papers/dig-lib-scope.html Requested on July 29, 2003.

[Lesk95] Lesk, Michael: “Why digital libraries?” June 19, 1995. http://www.ukoln.ac.uk/services/papers/follett/lesk/paper.html Requested on January 13, 2004.

[LoZv01] Loginov B.P., Zvetkova I.B. “Library standards and standardisation in the field of informational and searching systems” Russian Digital Libraries Journal - 2001 - Vol. 4 - No 4. http://www.iis.ru/el-bib/2001/200104/LC/LC.ru.html Requested on January 28, 2004.

[LyGM95] Lynch, Clifford; Garcia-Molina Hector:” Interoperability, Scaling, and the Digital Libraries Research Agenda”, August 22, 1995 http://www-diglib.stanford.edu/diglib/pub/reports/iita-dlw/ Requested on January 30, 2004.

[Mill00] Miller Paul: “Interoperability: what is it and why should I want it?”, June 21 2000 in Ariadne Issue 24. http://www.ariadne.ac.uk/issue24/interoperability/intro.html. Requested on January 17, 2004.

[NWDH96] Nikolaou, Christos; Wellman, Mike; Dallas, P.; Costis; Hugenholtz, Bernt (EU-NSF Working Group on Intellectual Property and Economic Issues): “Intellectual Property and Economic Issues for Digital Libraries: A Framework for Future Research. Report from the DLI workshop at Columbia University”, September 24-26, 1996. http://www.iei.pi.cnr.it/DELOS/REPORTS/ipe.htm

[Pask01] Paskin Norman: “DOI overview”. August 1, 2001. http://www.doi.org/overview/sys_overview_021601.html. Requested on January 12, 2004.

[PaWr02] Pasquinelli, Art; Wright, Cheryl D. Sun Microsystems, Inc.: “Digital Library Technology Trends”. August 2002. http://www.sun.com/products-n-solutions/edu/whitepapers/pdf/digital_library_trends.pdf . Requested on January 12, 2004.

[Pede00] Pederson, Gordon: “Models for Library Co-operation and the Need for Supporting IT Facilities”, 2000. http://www.iis.ru/el-bib/2000/200004/pedersen/pedersen.en.html Requested on January 23, 2004.

[Perr02] Perry, Rodney: “From institutional to production models of collaboration ” in First Monday, Volume 7, Number 5 (May 2002). http://firstmonday.org/issues/issue7_5/perry/index.html Requested on March 15, 2004.

[PlDi96] Place, Thomas W.; Dijkstra, Joost: “Z39.50 or WWW: which way to go?” in International Conference on Electronic Library and Visual Information Research 3. March 1996.

[Polo00] Polok, Margarete: „Strategien und Konzepte zur Langzeitsicherung digitaler Publikationen in wissenschaftlichen Bibliotheken“, 2000, in „Die neue Seite der Bibliothek“, 2001.

[PShJ96] Polger Thomas, Shapiro Celia D., Josephs Michael R.: “The concept of models of use and its application in digital libraries” in “International Conference on Electronic Library and Visual Information Research <3, 1996> : ELVIRA 3 : papers from the third ELVIRA conference”. March 1996.

[Raja02] Rajashekar, T.B.: “Digital Library and Information Services in Enterprises: Their Development and Management”. Year 2002. http://144.16.72.189/is214/214-2001-2002/topic-1.htm . Requested on January 10, 2004.

[RiWa98] Richvalsky James, Watkins David: “Design and Implementation of a Digital Library” in ACM Crossroads Student Magazine 1998.

http://www.acm.org/crossroads/xrds5-2/diglib.html. Requested on July 30, 2003.

[Sade01] Sadeh Tamar: "SFX and the OpenURL Framework for the Hybrid Library" June 6-8, 2001 http://www.stk.cz/elag2001/Papers/TamarSadeh/TamarSadeh.pdf . Requested on January 15, 2004.

[Scha95] Schatz Bruce: "Report of the Library Perspective Working Group. IITA Digital Libraries Workshop", May 18-19, 1995. http://www-diglib.stanford.edu/diglib/pub/reports/iita-dlw/part3.html. Requested on January 2, 2004.

[SHBA00] Van de Sompel Herbert, Hochstenbach Patrick, Beit-Arie Oren: „OpenUrl Syntax Description", June 16, 2000. http://www.sfxit.com/openurl/openurl.html. Requested on January 14, 2004.

[Sieg04] Siegel, Vivien: "The Public Library of Science".

http://www.publiclibraryofscience.org/.

Requested on August 12, 2004.

[Simm97] Simmons, Mike: "Collaboration dreams". http://www.dlib.org/dlib/march97/03editorial.html

Requested on November 6, 2004

[Simp97] Simpson, Donald B.: "Solving the challenges presented by electronic resources: creating opportunities through inter-institutional collaboration", 1997. In: Lee, Sul H.: "Economics of digital information", 1997.

[SMWB99] Sonnenwald, Diane H.; Marchionini, Gary; Wildemuth, Barbara M.; Dempsey, Bert J.; Viles, Charles L.; Tibbo, Helen R. "Collaboration Services in a Participatory Digital Library: An Emerging Design". 26 May 1999

http://ils.unc.edu/ils/research/reports/TR-2001-03.pdf

[TwNi98] Twidale, Michael; Nichols, David. “Survey of Applications of CSCW for Digital Libraries”, April 1998. http://www.comp.lancs.ac.uk/computing/research/cseg/projects/ariadne/docs/survey.html

[U.A.98] Unknown author: “Digital Library Definition for DLI2”. June 5, 1998. http://borg.lib.vt.edu/DLI2/defineDL.html . Requested on July 30, 2003

[U.A.03a] Unknown author. “The British library – about us”. http://www.bl.uk/about/cooperation.html Requested on September 2, 2003.

[U.A.03b] Unknown author. “About The European Library”. http://www.europeanlibrary.org/ Requested on July 2, 2003.

[U.A.03c] Unknown author. http://www.africandl.org/ Requested on September 4, 2003.

[U.A.04a] Unknown author, in Word Net Lexical Database, http://www.oxford-english.com/learn/WordNet/index.cfm. Requested on January 24, 2004.

[U.A.04b] Unknown author: “METS: An Overview & Tutorial” http://www.loc.gov/standards/mets/. Requested on January 14, 2004.

[U.A.04c] Unknown author. “Digital Library Initiatives”. http://memory.loc.gov/ammem/dli2/ Requested on March 3, 2004.

[U.A.04d] Unknown author. “Digital Cultural Heritage Community“. http://images.library.uiuc.edu/projects/DCHC/about.htm# Requested on March 23, 2004.

[U.A.04e] Unknown author. “About ED” http://www.ed.gov/about/landing.jhtml?src=ln Requested on March 29, 2004

[U.A.04f] Unknown author. About APIS. http://www.columbia.edu/cu/lweb/projects/digital/apis/about.html Requested on March 12, 2004

[U.A.04g] Unknown author. http://www.dli2.nsf.gov/internationalprojects/sino/CMNet.pdf Requested on March 13, 2004.

[U.A.04h] Unknown author. “ARIADNE Project Description”.

http://www.comp.lancs.ac.uk/computing/research/cseg/projects/ariadne/ariadne.html

Requested on November 6, 2004.

[U.A.04i] Unknown author. “Background”.

library.caltech.edu/openurl/Documents/Background_Feb_1st.doc

Requested on November 18, 2004.

[Visl01] Vislyi A.I.: “Digital libraries” Russian Digital Libraries Journal - 2001 - Vol. 4 - No 1. http://www.iis.ru/el-bib/2001/200101/vislii/vislii.ru.html. Requested on September 2, 2003.

[Wies00] Wiesenmueller, Heidrun: „Das Konzept der „Virtuellen Bibliothek“ im deutschen Bibliothekwesen der 1990er Jahre“. Year 2000.

[WRGr99] Wiseman, Norman; Rusbridge, Chris; Griffin, Stephen M.. “The Joint NSF/JISC International Digital Libraries Initiative” D-Lib Magazine Volume 5 Number 6 June 1999. http://www.dlib.org/dlib/june99/06wiseman.html Requested on February 3, 2004

[Zabe01] Zabelina, N.A.: **“**Corporate network Project- as information step for Networks of public libraries of Moscow”. Russian Digital Libraries Journal - 2001 - Vol. 4 - No 1. http://www.iis.ru/el-bib/2001/200101/zabelina/zabelina.en.html. Requested on January 27, 2004.

[ZnCh02] Znamenskiy A.V., Cherkalin S.D. “Integration of regional museum information resources” in Russian Digital Libraries Journal - 2001 - Vol. 4 - No 5. http://www.iis.ru/el-bib/2001/200105/ZCh/ZCh.ru.html. Requested on January 28, 2004.