C O L D L i b
A concept for a global knowledge space for the COLLNET community

Bernd Markscheffel[1]   Hendrik Thomas[2]

02 June 2008

Abstract

Main objective of the COLDLib project is the development of a global knowledge space for the interdisciplinary research network "Collaboration in Science and in Technology" - COLLNET to collect, combine and enrich the existing knowledge in the field of quantitative aspects of science of science, collaboration and communication in science and technology, science policy, combination and integration of qualitative and quantitative approaches, issues in scientometrics/ informetrics/ webometrics. With the help of this Digital Library, it will be possible to search information, which are currently worldwide scattered in various silos, in different software applications, different languages, heterogeneous data formats and different places in a semantically interconnected way.

1   Introduction

Kretschmer, H., Liang, L. and Kundra, R. (2001), 531-537 mentioned a worldwide increasing trend in all forms of cooperation and collaboration in science and technology. This trend is often described in bibliometric- and scientometric terms, mostly with international co-authorships as indicator. It is closely connected with a so-called "new mode of knowledge production" by which the sudden emergence of new combinations of formerly disconnected research areas, even from quite different disciplines, is meant.
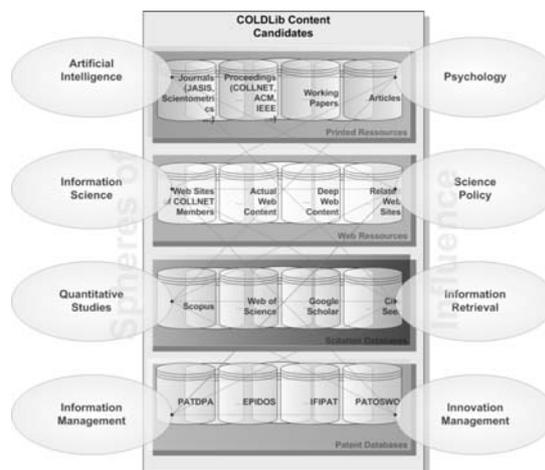

Figure 1: Potential Content Sources and Spheres of Influence of COLDLib

To improve the research on this highly important topic, a global interdisciplinary research network under the label "Collaboration in Sci-

---

[1] Chair of Information- and Knowledge Management,
   Technische Universität Ilmenau,
   P.O. Box 100565, 98693 Ilmenau, Germany
   bernd dot markscheffel at tu minus ilmenau dot de
[2] s. footnote before, hendrik dot thomas at tu minus ilmenau dot de

ence and in Technology" was founded 2000 in Berlin. COLLNET is an international network of investigators which aim for a deeper understanding of collaborative activities and processes.

Figure 1 shows the various spheres of influence in this complex research area and the network of implications between them. However, the worldwide existing knowledge about collaboration in theory and practice is mostly scattered and not always accessible. This does not comply with today's requirements concerning an effective information retrieval.

2　What is a Digital Library?

A generally accepted definition of the term digital library has not yet emerged (Leiner, B.M. (1998)). The diversity of definitions in literature shows that it is difficult to define this term clearly. Besides the term digital library other frequently used terms are electronic library, virtual library, and e-library. Many authors use these terms synonymously (Waters, D.J.H. (1998)). Other authors make a distinction and refer to virtual libraries as ordered link collections for selected information sources (Ewert, G. and Umstaetter, W. (1997)).

Contrary to traditional libraries, virtual libraries do not store information sources. Instead, they organize subjects and interconnect them to relevant information sources. An electronic library is a library that offers digital or digitalized information sources and computer aided information services. Umstaetter, W. (1995) combined the terms "virtual" and "electronic" library to the new expression "digital library". Another quoted definition was published by Lesk, M. (1997). He describes digital libraries as:

> „...*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*".

Other authors do not define the term explicitly, but list essential definition elements.

- "The digital library is not a single entity;
- The digital library requires technology to link the resources of many;
- The linkages between the many digital libraries and information services are transparent to the end users;
- Universal access (anytime and anywhere) to digital libraries and information services is a goal;
- Digital library collections are not limited to document surrogates: they extend to digital artefacts that cannot be represented or distributed in printed formats" (Drabenstott, K.M. (1994)).

Waters, D.J.H. (1998) combines various definitions and develops a more comprehensive definition. He describes digital libraries as, „organizations that provide resources including the specialized staff to select, structure, offer intellectual access to, interpret, distribute, preserve the integrity of, and ensure the persistence over time of collections of digital works so that they are readily and economically available for use by a defined community or set of communities".

The DELOS (2005) -project extends this definition. According to its vision, the New Generation Digital Libraries:

> "..*should not just be seen as static information repositories but as growing, inter-actively, and collaboratively used nuclei of what will be at some stage, a good part of human knowledge that depends as much on information as on communication.*"

Based on these definitions and the discussion of the term information services we propose the following definition of digital libraries:

They are structured electronic collections of documents. Digital libraries comprise documents in heterogeneous media form; e. g. text, structured data, pictures, audio, video, and executable software. Digital libraries are accessible via internet technologies or they are available on CD-ROM, DVD, or other media. They support activities to satisfy information needs and help to solve specific information problems. Digital libraries perform a wide range of different activities. In the most basic form, digital libraries transfer traditional services of classic libraries (structuring, classification und provision) to the digital world. More sophisticated business models provide additional services (e.g. collaboration or e-learning services). Specialized business models provide particular

information services across multiple digital libraries (e.g. meta classification schemes or searching facilities). We refer to all these business models as digital libraries. Our definition includes New Generation Digital Libraries in the sense of the DELOS-project. Digital libraries are a specific form of information services.

3   COLDLib Approach

3.1   Previous approaches

Havemann, F. and Katz, J.S. (2004) mentioned the necessity for an effective collaboration for the COLLNET community first. They called the concept CollBib and declared the following main features:

- Complete bibliographic information including article title, journal, volume, page number, abstract, names and addresses of all the authors listed on the paper,
- Credits to the COLLNET members who entered the bibliographic information,
- A facility for members to add comments to individual bibliographic records,
- A means for COLLNET members to correct a bibliographic entry,
- A convenient way to download bibliographic records from CollBib to a PC database.

Another community project has been proposed by Roland Wagner-Döbler. The main idea behind the COLLSpace approach was to establish a virtual Workspace where the COLLNET community members could discuss new projects and share their personal data by using an email reflector and a central document repository which is accessible by a web browser.

These are useful ideas for the COLLNET community to build an open collaborative platform. However, till today none of these promising concepts has been implemented or generally approved in the COLLNET community. Concluding a suitable information space for COLLNET is still missing, which is quite odd, given the acknowledged importance of collaboration and information access for the research in Science and in Technology.

3.2   DMG-Lib

The project "Digital Mechanism and Gear Library" (DMG-Lib) was started as an interdisciplinary project of the Technical Universities of Ilmenau, Aachen and Dresden in 2004. It is founded by the "German Research Foundation" in the program "Scientific Library Services and Information Systems" (project number: LIS 4-554975). The objective of this project is the development of a digital, internet-based library which contains great amounts of the worldwide knowledge about mechanisms and gears. In contrast to other Digital Library projects the digitalisation of all available and accessible resources is not the final aim of the DMG-Lib activities. To offer users a wide variety of opportunities for retrieval, the digitized resources are enriched by additional multiple information like animations, references, and explanations. The library is designed to satisfy the requirements of different user groups like engineers, librarians, teachers, students, historians and the interested public.

The DMG-Lib online portal on www.dmg-lib.org contains a vast amount of digital documents representing more than 1000 gear models, 100 machines, 3.500 photographs, 100 videos and animations, 400 books published before 1898 and nearly 10.000 "classical" documents, such as technical reports, patents, research papers and books. These digital documents are represented in various formats and media. DMG-Lib is developed by approx. 30 engineers at 8 different locations.

A universal solution to the problem of structuring of such a large information space demands an interdisciplinary, distributed, and collaborative approach[3].

---

[3] TMwiki (Markscheffel, B. Thomas, H. and Stelzer, D. (2006)) is a first step towards building such a collaborative environment. It also provides a fundamental part for a sophisticated navigation tool for the DMGLib-Portal. It is a web application that allows users to add and edit content. TMwiki is based on DokuWiki (Gohr, A. (2006)), a standards compliant and simple to use Wiki engine.

TMwiki supports collaborative work by providing numerous features. Besides standard Wiki-features such as collections of relevant resources (Topic Map software, Topic Map samples, glossary, web pages, blogs), mailing lists and discussion areas, RSS feeds and a tutorial section, it also provides access to some helpful Topic Map tools, e.g. MERLINO (a tool for semi automated generation of occur-

In the next years thousands of resources will be provided. This implies a key challenge of this project: the implementation of an efficient, uniform and user-satisfying information retrieval over the huge amount of heterogeneous information resources. The necessary new level of quality for the information retrieval will be achieved by means of Topic Maps. A semantic meta-layer is used to integrate different terminologies, languages, research paradigms and historic eras in which digital documents were developed. This facilitates information retrieval in voluminous and complex libraries.

These described requirements for information retrieval and management are very similar to the needs of the COLLNET community. So it is reasonable, that our approach for COLDLib is greatly influenced by our work in the DMGLib project.

3.3 COLDLib concept description

The main objective of this digital library project for the COLDLib can be derived from the tasks of traditional libraries:

- Collecting
- Systematisation
- Preserving and
- Providing

of information from and for the COLLNET community.

But, moreover it can serve as an integration, discussion and collaboration platform for delegates of the in Fig. 1 shown spheres of influence. The term integration can be interpreted on the content level. That means, with the use of a topic map based semantic meta-layer it will be possible to interconnect the various resources on an ontology-like as well as on a subject-centric level. So, the retrieval ability of the digital library is more suitable to the variety of information needs in this heterogeneous community.

On the other hand the term integration can also be interpreted on the collaboration level in terms of an efficient communication and knowledge exchange between the involved interdisciplinary partners.

Based on these requirements we developed a workflow and corresponding working packages for the COLDLib, which is shown Figure 2.

At the first glance the separation of the portal- and the production database is obviously. The production database can serve as a cache for the several enrichment steps of the raw data. Users query only the portal database, which has the enriched content within.
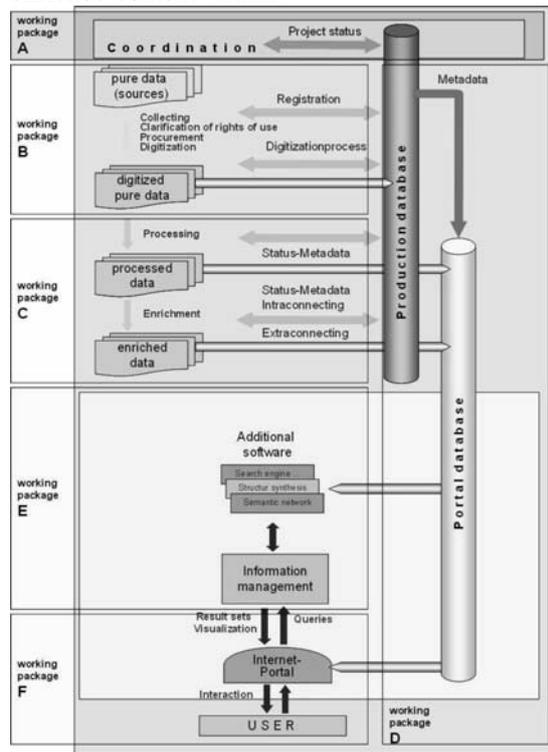

Fig. 2: Workflow and Working Packages of COLDLib[4]

The several working packages describe the workflow in a more task distributed than workflow oriented matter.

Package A deals with the necessary project management processes, e.g. project coordination, - organisation and – promotion (participation on conferences, exhibitions, fairs…).

Package B comprises the information resource management oriented tasks. In this package the potential information resources must be identified, collected and - if necessary - digitized, the legal rights of use have to be clarified

---

rences in topic maps. TMwiki also offers a sandbox for exploring and testing these applications.

[4] Workflow and working packages are adopted from the DMGLib project (Trott, Döhring, Brix and Brecht 2007). It can serve as a model for the COLDLib.

and the amount of content must be transferred to the production database.

Package C contains features which extended the functionality of traditional digital libraries. It includes the post-processing of the raw data (converting of digitized raw data, converting of pictures, analyse and converting of video and audio files), the content selection for the enrichment and the enrichment itself (enrichment with meta-data, extra-connection with actual web content, intra-connection with other stored topics based on the semantic meta-layer).

Package D covers the more technical problems of COLDLib, mainly the operation and maintenance of the COLDLib server and the creation and evaluation of usage staticstics.

Package E is dealing with the key challenge of such a heterogeneous information resources with different levels of complexity – the provision of effective information retrieval mechanisms. Semantic Technologies, e.g. Topic Maps, Frames or semantic markup languages, can be used to structure heterogeneous information resources. Thus, the complex relationships of information resources (see package C) can be modelled with the help of semantic technologies. These technologies and first of all Topic Maps can not only replicate the features of those traditional library science techniques, they can go far beyond (Hyun-Sil, L., Yang-Seung, J., Sung-Kook, H. (2005), Pepper, S. (2000), Schweiger, R., Dudeck, J. (2005)). Topic Maps as a semantic meta-layer can be used to model the structure of knowledge implicit available in the information resources (Markscheffel, B., Thomas, H. and Stelzer, D. (2005)). Information retrieval concepts and systems driven by such rich semantic information can support a more precise and flexible information retrieval and structuring of the heterogeneous information resources.

Package F includes additional technical features like the portal construction, -operation and maintenance and also the usability engineering of the GUI.

Package G which is not shown in figure 2 focuses on the tasks for realizing a sustainability, like the analysis of concepts for a suitable business model (Markscheffel, B. Fischer, D. (2007a,b) Markscheffel, B, Fischer, D. and Stel-

zer, D (2008)) and the analysis of financial capabilities.

4 Conclusion

This paper should be interpreted as a proposal for a project draft. Based on the discussed requirements and proposed features hopefully a fruitful discussion in the COLLNET community can emerge. The future steps should be:

- Identify suitable funding programs to support and finance the project,
- Identify a set of competent partners, which are able to collaborate in this task,
- Assign responsibilities for the different working packages to the involved partners,
- Define a project plan for the realisation of the COLDLib project.

With the help of these ideas it should be possible to build a global knowledge space which provides an easy to use and efficient access to the knowledge of COLLNET community in order to enhance the usage of the gained knowledge and provide a holistic and federated instead of the current isolated view on the information.

References

DELOS (2005). *About DELOS*. http://www.delos.info/index.php

Drabenstott,K.M. (1994). *Analytical review of the library of the future - Council Library Resources*. Washington DC.

Ewert, G. and Umstaetter, W. (1997). *Lehrbuch der Bibliotheksverwaltung*. Stuttgart: Hiersemann Verlag.

Gohr, A. (2006). *Wiki:dokuwiki*. http://wiki.splitbrain.org/wiki:dokuwiki

Havemann, F. and Katz, J.S. (2004). *COLLNET and Virtual Collaboration*, http://bscw.gmd.de/pub/english.cgi/0/28488 300/COLLNET%20and%20Virtual%20Col laboration.rtf

Hyun-Sil, L., Yang-Seung, J., Sung-Kook, H. (2005). MARCXTM: Topic Maps Modeling of MARC Bibliographic Information.

In: Maicher, L. and Park J. (Eds.): *Charting the Topic Maps Research and Application Landscape*. Springer, Berlin, pp. 240–250.

Kretschmer, H., Liang, L. and Kundra, R. (2001). Foundation of a Global Interdisciplinary Research Network (COLLNET) with Berlin as the Virtual Centre. In: *Scientometrics*, Kluwer Academic Publishers, Dordrecht Vol. 52, No. 3 (2001) pp. 531–537.

Leiner, B. M. (1998). *The Scope of the Digital Library. Draft for the D-Lib Working Group on Digital Library Metric,* http://www.dlib.org/metrics/public/papers/dig-lib-scope.html

Lesk, M. (1997). *Practical digital libraries: Books, bytes and bucks*. San Francisco: Morgan Kaufmann.

Markscheffel, B. Fischer, D. (2007a). Business Model Development for Digital Libraries. In*: Online Information 2007 Conference Proceedings (Online Version)*, London 04.-06.12.2007, London, pp. 39-41.

Markscheffel, B. Fischer, D. (2007b). A Roadmap for Developing a Revenue Model for Digital Libraries - a Case Study for the Digital Mechanism and Gear Library. In: Youakim Badr, Richard Chbeir, Pit Pichappan (Eds.): *Proceedings of the Second IEEE International Conference on Digital Information Management (ICDIM 2007)*, Lyon 28.-31.10.2007. Los Alamitos, Washington, Brussels, Tokyo 2007, pp. 61-66.

Markscheffel, B. Fischer, D. and Stelzer, D. (2008). Classification of Digital Libraries - An e-Business Model-Based Approach. In: *Journal of Digital Information Management*. Vol. 6, Nr. 1, 2008, pp. 71-80.

Markscheffel, B., Thomas, H. and Stelzer, D. (2005). Merlino - a prototype for semi automated generation of occurrences in topic maps using internet search engines. In: *Posters and Demos, 2nd European Semantic Web Conference 2005*, Heraklion, Greece, 29..-01.06.05, pp. 25 -27.

Markscheffel, B., Thomas, H. and Stelzer, D. (2006). TMwiki - a Collaborative Environment for Topic Map Development. In:

Demo and Poster Proceedings of the 3nd European Semantic Web Conference (ESWC 2006), Budva, Montenegro, 11.-14.06.2006.

Pepper, S. (2000). *The TAO of Topic Maps.* http://www.gca.org/papers/xmleurope2000/papers/s11-01.html

Schweiger, R., Dudeck, J. (2005). Improving Information Retrieval Using XML and Topic Maps. In: Maicher, L., Park, J. (Eds.): *Charting the Topic Maps Research and Application Landscape*. Springer, Berlin, pp. 251–263.

Thomas, H., Markscheffel, B. and Brix, T. (2006). *SIREN - a Topic Map based Semantic Information Retrieval Environment for Digital Libraries*. Poster-Presentation TMRA 2006, Leipzig, Germany 11.-12.10.2006.

Trott, S.; Döring, U.; Brix, T. and Brecht, R. (2007). Die Digitale Mechanismen- und Getriebebibliothek "DMG-Lib" als Modell für die Integration heterogener Informationsquellen. In: Stempfhuber, M. (Ed.): *Lokal - Global: Vernetzung wissenschaftlicher Infrastrukturen. 12. Kongress der IuK-Initiative der Wissenschaftlichen Fachgesellschaft in Deutschland* (28.09.2006 in Göttingen) - Bonn: GESIS - IZ Sozialwissenschaften 2007 (Tagungsberichte).

Umstaetter, W. (1995). Die Rolle der Dokumentation bei der Entstehung der Digitalen Bibliothek und ihre Konsequenzen für die Bibliothekswissenschaft. *In: Nachrichten für Dokumentation,* vol. 46, 1995, pp. 33-42.

Waters, D.J.H. (1998). *What Are Digital Libraries?* CLIR issues, 4 ed.