

CASLIN 2009

16th International Seminar

Institutional Online Repositories and Open Access

Teplá Monastery, Czech Republic
7-11 June 2009

CASLIN 2009

Institutional Online Repositories and Open Access

16th International Seminar

Teplá Monastery, Czech Republic, 7-11 June 2009



Pilsen 2009

University of West Bohemia

Committees and Organizers

Program Committee

Martin Svoboda, *State Technical Library, Prague, Czech Republic*

Andrew Lass, *Mount Holyoke College, South Hadley, MA, USA*

Miloslava Faitová, *University of West Bohemia, Pilsen, Czech Republic*

Dana Lošťáková, *Palacký University, Olomouc, Czech Republic*

Ivana Horáková, *Study and Research Library, Pilsen, Czech Republic*

Alojz Androvič, *University Library in Bratislava, Slovakia*

Petr Žabička, *Moravian Library, Brno, Czech Republic*

Martin Lhoták, *Czech Academy of Sciences Library, Prague, Czech Republic*

Miroslav Bartošek, *Masaryk University, Brno, Czech Republic*

Organizing Committee

Miloslava Faitová, *University of West Bohemia, Pilsen, Czech Republic*

Irena Šléglová, *Masaryk University, Brno, Czech Republic*

Barbora Katolická, *University of West Bohemia, Pilsen, Czech Republic*

Radka Tichá, *University of West Bohemia, Pilsen, Czech Republic*

Zdenka Firstová, *University of West Bohemia, Pilsen, Czech Republic*

Aleš Horák, *Palacký University, Olomouc, Czech Republic*



ISBN 978-80-7043-806-0

Published by University of West Bohemia, Pilsen

Johan van HALM

* 1941 – † 2008

International consultant and friend of the CASLIN project



Previous CASLIN Conferences

CASLIN 2008

Theme: Knowledge Technologies and (the Future of) Higher Education

Date and place: Blansko, 15. – 19. 6. 2008, Czech Republic. Organised by Masarykova univerzita v Brně

CASLIN 2007

Theme: IT & scholarly communication

Date and place: Stupava, 3. – 7. 6. 2007, Slovakia. Organised by University Library Bratislava

CASLIN 2006

Theme: Feedback: How well do we do it?

Date and place: Český ráj, 11. – 15. 6. 2006, Czech Republic. Organised by Library of the Academy of Sciences of the Czech Republic

CASLIN 2005

Theme: Public Services: New Trends

Date and place: 19. - 23. june 2005 in Lednici na Moravě, Czech Republic. Organised by Moravian Library in Brno

CASLIN 2004

Theme: The Management of Memory: The Gray Areas

Date and place: 5 - 9 September 2004, Čingov - Slovenský raj, Slovakia. Organised by KOLIN Consortium

CASLIN 2003

Theme: Crisis Management and Recovery

Date and place: 8 - 12 June 2003, Holenský dvůr u Jindřichova Hradce, Czech Republic.

Organised by State Technical and National Library of Czechia, Prague

CASLIN 2002

Theme: Document Preservation and Access: New Trends

Date and place: 23 - 27 June 2002, Podbanské, Slovakia. Organised by Slovak National Library in co-operation with CASLIN

CASLIN 2001

Theme: Document Description and Access : New Challenge

Date and place: 27 - 31 May 2001, Beroun, Czech Republic.

Organised by Library of Academy of Sciences of Czech Republic and National Library of Czech Republic in co-operation with CASLIN

CASLIN 2000

Theme: Education for Librarians

Date and place: 28 May - 1 June 1998, Luhačovice, Czech Republic.

Organised by MOLIN Consortium and National Library of Czech Republic

CASLIN'99

Theme: Union Catalogs: Their Organization and Service

Date and place: 30 May - 3 June 1999, Zadov na Šumavě, Czech Republic.

Organised by State Technical Library, National Library of Czech Republic, Masaryk University

CASLIN'98

Theme: Libraries Without Walls

Date and place: 31 May - 4 June 1998, Štrbské Pleso, Slovakia. Organised by KOLIN Consortium

CASLIN'97

Theme: Libraries of the Future

Date and place: 8 - 12 June 1997, Dlouhé Stráně, Czech Republic. Organised by MOLIN Consortium

CASLIN'95

Theme: Library Consortia

Date and place: 27 - 31 August 1995, Smolenice, Slovakia. Organised by University Library Bratislava

CASLIN'94

Theme: Management Training

Date and place: 19 - 23 June 1994, Liptovský Ján, Slovakia. Organised by Matica slovenská - Slovak National Library, Martin

CASLIN'93

Theme: Management Training

Date and place: 5 - 9 September 1993, Rožnov pod Radhoštěm, Czech Republic. Organised by The Moravian Library, Brno

CASLIN 2009: Institutional Online Repositories and Open Access

Preface

While research and university libraries increasingly serve as on-line providers of searchable, full text materials, the institutions that they serve are focusing their attention on using these portals as platforms for their own publishing to a broader even global community of users. These institutional on-line repositories, and the idea of Open Access, mark another significant development in how knowledge is produced and made available. It also offers new possibilities for change and some significant challenges as well to the libraries themselves.

CASLIN 2009 will, as is our custom, offer the participant the chance to hear from international experts who are working in this area, as well as a platform to discuss their own experience.

The Tertiary Educational Sector of New Zealand provides the following useful definitions: "An Institutional Repository (IR) is a set of services for storing and making available digital research materials created by an institution and its community — digital collection of the community's research output.

IRs form part of a larger national, regional and global system of Open Access repositories, indexed in a standardized way and searchable using one

interface. They can also be the foundation for new models of scholarly publishing and communication. They are an essential tool for scholarly communication in the digital age. They can be linked to related services such as eLearning, content preservation and migration, and research management and reporting."

"The Open Access movement aims to transform scholarly communication by making it easier for researchers to find and share the results of research, through free and unrestricted on-line availability. The open access research literature consists of on-line copies of peer-reviewed journal articles, conference papers, theses, technical reports and working papers. In most cases there are no licensing restrictions on their use by readers. They can be used freely for research, teaching and other purposes."

"Open access is not self-publishing, nor a way to bypass peer-review and publication, nor is it a low quality alternative publishing channel. It is simply a way to make copies of research results freely available on-line to the whole research community."

Andrew Lass, Mount Holyoke College, USA

Contents

Exploring Ways that Institutional Repositories Facilitate New Roles and Partnerships for Libraries and the Academy Marilyn S Billings (University of Massachusetts, USA)	7
The End of Institutional Repositories & the Beginning of Social Academic Research Service: an Enhanced Role for Libraries Stuart Basefsky (Cornell University, USA).....	15
A Short Overview of International Trends in Open Access Leo Waaijers (Open Access consultant, Netherlands).....	23
Open Access and Repositories: A Status Report from the World of High-Energy Physics Jens Vigen (European Organization for Nuclear Research, CERN, Switzerland).....	34
Open Acces and Institutional Repositories from Local Initiatives to Global Solutions Friedrich Summann (Bielefeld University Library, Germany)	39
Users' Behavior and Institutional Repositories: Towards Information Ecology Jela Steinerová (Comenius University Bratislava, Faculty of Philosophy, Slovakia).....	43
Institutional Repositories and Document Citation Linda Skolková & Miloslav Nič (Institute of Information Studies and Librarianship, Faculty of Arts, Charles University, Prague, Czech Republic / Institute of Chemical Technology, Prague, Czech Republic)	50
VirtuReal Open Access 2.0 Jakub Štogr (Charles University in Prague, Czech Republic)	61
Open Access Institutional Repository and the Role of Academic Library Staff: Reconsidering the Position of the Academic Library in Preserving and Disseminating University's Scholarly Output Daniela Tkačíková (VŠB-Technical University of Ostrava, Czech Republic)	65
Open Access and Institutional Repository: Approach of the Czech Academy of Sciences and Link up to International Projects Martin Lhoták & Roman Chýla (Czech Academy of Sciences Library, Czech Republic)	71
An Experience with Building Digital Open Access Repository DML-CZ Petr Sojka (Faculty of Informatics, Masaryk University Brno, Czech Republic)	74
Open Source versus Commercial Solutions for a Long-term Preservation in Digital Repositories Andrea Fojtů (Charles University of Prague, Czech Republic)	79
National Repositories of ETDs and Grey Literature in the Czech Republic Iva Horová & Jan Mach (Academy of Performing Arts in Prague, Library, Czech Republic / University of Economics, Prague, CILS, Czech Republic)	80
Institutional Repository at the University Iva Prochásková & Lucie Vyčítalová (University of Pardubice, Czech Republic)	81

Exploring Ways that Institutional Repositories Facilitate New Roles and Partnerships for Libraries and the Academy

Marilyn S Billings

mbillings@library.umass.edu

University of Massachusetts, USA

Abstract: Libraries have been functioning in the digital publishing arena for a few years now. How has this new service affected the way that we envision our place in the academy? What new tools are we using and skills are we developing to fulfill this service? How has the academy changed its view of the library's role in the university context? What new partnerships have we created and fostered to exploit this new vision? This presentation will foster discussion of these and similar questions, providing insights and opportunities for further exploration of how the role of libraries as publishers enables us to be key partners in the creation, dissemination, and archiving of academic scholarship.

Keywords: institutional repositories, library partnerships, digital publishing, open access, new roles for librarians

Background

The University of Massachusetts Amherst (UMass Amherst), the flagship campus of the University of Massachusetts system, sits on nearly 1,450-acres in the scenic Pioneer Valley of Western Massachusetts, 90 miles from Boston and 175 miles from New York City. The campus provides a rich cultural environment in a rural setting close to major urban centers. It has 26,360 total undergraduate and graduate students and 1,180 full-time instructional faculty.

The UMass Amherst Libraries is a system consisting of the W.E.B. Du Bois Library (including the Learning Commons that Jay Schafer discussed at last year's conference) and the Integrated Sciences and Engineering Library. The system also manages the Image Collection Library and the Music Reserve Lab. The Scholarly Communication Office is located on the 16th floor of the W.E.B. Du Bois Library and is responsible for teaching librarians, faculty, and students about new models for scholarly communication to be discussed later in this paper and for managing the digital repository, ScholarWorks @ UMass Amherst.

Scholarly Communication Crisis

A major driving force behind the development of institutional repositories has been the dramatic shift in scholarly communication, especially since the early 1990s. This shift has been caused primarily by two factors, the increasing volume of information in digital form and spiraling publishing costs. Traditional print publications (books, journals) are being augmented by, in some cases replaced by, materials born digital. Spiraling costs of journals have necessitated major cuts in journal

subscriptions, resulting in fewer journal purchases and reduced access to information. It has become clear that the changing economics of scholarly publishing are proving to be much too costly to sustain in the long run. We need new models for managing our scholarly research output and for disseminating that information to our users. Some of the issues we need to address when creating the new model include what we collect, how we address the needs of researchers, and how we preserve and make accessible this body of material. This evolution of scholarly communication, using new technologies, digital information, and networked environments that are prevalent on college and university campuses today, has created a truly new environment for scholarly research which requires broad campus conversation.

Development of an IR at UMass Amherst: a case study approach

In the fall of 2003, I submitted a sabbatical leave proposal to investigate the current environment of institutional repositories to determine if such a project would be feasible for UMass Amherst to undertake. The University Libraries awarded the sabbatical research leave for the first half of 2005. We will examine the findings from that sabbatical and use the development of the IR as a case study. First we need to define what we mean by an IR and there is an excellent definition in our program materials. IRs are seen as one of the key solutions to the scholarly communication crisis just mentioned and to resolve the advent of a variety of independent, silo-like web sites that store and provide access to digital materials. IRs go beyond the technology. They foster collaboration among librarians, archivists, information technology

providers, the administration and others at the university to provide the institutional support and resources needed for the IR's success. Their quick rise addresses many needs that have been identified in the scholarly communication crisis: the escalating costs of scholarly materials, especially STM journals, the growing diversity of content in a variety of formats that need to be grabbed and preserved before the content has disappeared, escalating user expectations (immediacy, relevance, high level of management of content, resulting from experience with Google, Amazon), and faculty interest in new forms of scholarly communication for research and teaching.

Benefits of an IR

According to Kathleen Shearer in the Association of Research Libraries webcast¹ in June 2004, a centralized repository provides common formats, uniform, well-defined structures, provides other features such as searching or links from other sources to be applied consistently across the entire collection, data is more accessible and easier to use, greater integration with other related resources, easier to develop additional tools for more functionality. Institutional Repositories have a unique niche in the organization to collect, organize, and provide access to a wide range of content that was previously inaccessible or scattered. They are now an important step in the evolution of scholarly communication. In the same webcast, Daniel Greenstein from the California Digital Library described many value-added services that an IR can provide in supporting scholarship:

- Greater links between content in IRs and scholarly publications
- Better search and retrieval facilities
- Enhance existing scholarly publishing (pre-prints, etc in digital form and interlink with paper)
- Move scholarly communication forward
- Harvesting of research
- Broad range of scholarly materials
- Enhance professional visibility driven by broader dissemination and increased use

Core features

Common features of a repository repeated in numerous research articles during my research include the following elements

- content is institutionally defined and scholarly in nature
- cumulative and perpetual
- open and interoperable
- content is in digital form in a wide variety of types (text, audio, video, images, data sets)
- community focus, where the community determines what is included in the repository. The community members are the authors and copyright owners of the deposited content

¹ ARL/OLMS Webcast. *Institutional Repositories : Revealing Our Strengths*. Available online at <http://www.library.arizona.edu/documents/dlist/IR_final.pdf>.

- institutional support, requiring collaboration across an organization. The repository requires long-term financial support to ensure that the content is preserved and maintained
- durable (persistent url for material), permanent content that can be migrated over time. Digital preservation techniques need considerable more research to ensure lasting retrieval
- access to content by a broad audience, a community-shared alternative to local storage of content, fosters serendipitous discovery across disciplines
- Personalization of research pages by faculty
- Authoring needs such as version control, easy access to files

Core functionality

Core functions that were emphasized during my research site visits which are critical for a successful repository are included in the following list:

- material submission, some way for the author to deposit material, provide for editing to assure quality of content, conversion to archival format such as PDF
- metadata application, such as author, title and descriptive information and administrative data such as date and time of submission
- access control, or digital rights management, to provide for controlled access to the repository content. Even if the entire community has access to the content, there needs to be a way to restrict the ability to add, delete, edit, and approve content.
- discovery support, usually a search engine that supports browsing and full-text searching of the content
- distribution and dissemination of content to enable display and download capabilities for further manipulation of content for ongoing research
- preservation, some mechanism(s) for the content to be preserved and retrievable over time, including a persistent documentation identification system
- batch loading capability
- indexing of content based on OAI metadata harvesting protocol.

Challenges / Issues for IRs:

Populating the Institutional Repository

There has been an acknowledgement among early adopters of institutional repositories that recruitment of faculty content has been a challenge that needs to be addressed in order for IRs to be successful. Frequently faculty and researchers are unaware of the potential benefits of an IR as outlined above. Recent research conducted at the University of Rochester (Gibbons and Foster), has resulted in some concrete methodologies for others to follow. Solutions include talking to faculty in their own language about grey literature, collaborating with the university press concerning monographs and faculty working papers or series, and batch loading of content

from publishers adhering to open access principles. When the IR is presented as infrastructure for “non-traditional” material rather than published literature, then there is tremendous interest by faculty.

Some examples of possible content include:

- Working papers and other “gray” literature”
- Peer reviewed series (author or organization may want to create online version) and platforms for peer-reviewed journals (place where these can be created)
- Platforms for producing and distributing monographs
- Data archives, large sets of research data
- Classroom teaching materials
- Digital archives
- Print repositories
- Student electronic portfolios
- Institution’s annual reports
- Computer programs
- Photographs
- Art works
- Pre-prints
- Peer-reviewed articles
- Conference papers
- Electronic theses and dissertations

Policy Development

A successful IR cannot be developed without giving serious consideration to its overall structure and design. One recommendation from early adopters is that an IR needs a precise definition and needs to have a clear understanding of what the IR will provide. This overall structure and governance can be initiated by developing policies that, during pilot and implementation phase, should be kept in draft form to provide structure but also be flexible enough to change easily until it is working the way you want it to operate. Policies can then be finalized once the IR is in general release.

Policies can be separated out into the following areas:

- Services
- Content acquisition and management
- Preservation and access
- Rights management

There are many resources for policy development that were developed by early IR adopters which are listed in the bibliography.

Funding / Business Models

Libraries have been struggling with how to fund the IR in a sustainable way. Recent research by Soo Young Rieh² indicates that there are a wide variety of ways, many informal, that this is being accomplished. A special initiative supported by the library, costs absorbed in routine

library operating costs, regular budget line item for your institution's library, grant awarded by an external source, special initiative supported by your institution's central administration, or a special initiative supported by the institution's archives are some models. There remains a lot of work to be done to determine best funding and business models however.

Legal Issues

The top two issues that need to be addressed are intellectual property issues and copyright. These issues are further complicated in the international academic and research environment. Creative Commons licenses <http://creativecommons.org/> provide **free** licenses and other legal tools to mark creative work with the freedom the creator wants it to carry, so others can share, remix, use commercially, or any combination thereof. The Czech Republic license has now been integrated into the Creative Commons licensing process.

<http://creativecommons.org/international/cz/>.

Preservation

This challenge remains a long way from being solved. There are at least four possible approaches to the problem: migration, technology preservation, emulation, and persistent object preservation. Libraries, especially in traditional special collections and archives areas, have expertise in this area that can be expanded and extended to IR development. Another way to put the IR in the best position for long-term preservation is to adhere to industry standards and best practices. IR managers will need to develop coordinated strategies to ensure long-term preservation of the IR contents. One international example to monitor is CLOCKSS which is a joint venture between the world's leading scholarly publishers and research libraries. Its mission is to build a sustainable, geographically distributed dark archive with which to ensure the long-term survival of Web-based scholarly publications for the benefit of the greater global research community.

Other areas that have continued to provide areas of challenge for IR development both at the time of and since my sabbatical research include marketing and publicity, organization and administration, and staffing which will be discussed next as part of the „Lessons Learned“ section.

Lessons learned

Since much of the sabbatical research included site visits or extensive phone calls to early adopters of IRs, we learned many valuable lessons and obtained a lot of guidance to help UMass Amherst avoid many of the potential problems that they had experienced. In addition, the contacts made at workshop gave us additional colleagues with whom to work as we developed our IR.

² RIEH S.Y. et al. *Census of Institutional Repositories in the U.S.: A Comparison Across the Institutions at Different Stages of IR Development*. <http://www.dlib.org/dlib/november07/rieh/11rieh.html>.

Lesson 1: Communications

- Start talking about scholarly communication and the benefits of a repository early in the process
- Educating the faculty is of key importance. They need to understand the importance of a repository and adjust customs and culture
- Repository policies need to have clear explanations and examples. Use policies, etc from early adopters³

Lesson 2: Planning

- Simultaneous with technical planning and installation, do business and operational planning
- Be flexible and responsive, especially when defining service model. Be prepared to change during implementation
- Prepare for additional storage capacity

Lesson 3: Recruitment of early adopters

- There's no such thing as too much publicity
- Starting new communities is a long process, don't get frustrated
- Community policies need to be established at the top level, for example, chair of an academic department, head of a research institute, etc
- Use terminology that faculty members understand. Visibility of their content and digital preservation are critical issues for them. Don't get technical
- Personal contact is the best means for finding pilot projects. Talk with individual faculty and small groups
- When considering pilot projects, select those that provide a diversity of formats, storage requirements, and size
- Collection development coordinators are most helpful in recruiting early adopters
- Survey content providers / communities yearly for use, management, etc. Stay in touch with them, provide training and other support as needed

Lesson 4: Provide Talking Points (Selling Points)

- Persistent identifiers for content; visibility of content
- One uniform presence for institutional research
- Community and author control
- Self-archiving clause for green publishers
- Digital preservation of approved formats
- Visibility of research leads to more recognition, contribution to scholarship in discipline

Lesson 5: Staffing

- Educate library staff through presentations, frequent updates in library communications, updates at departmental and staff meetings to increase awareness and understanding

- Use cross-functional library staff teams when building repository's business plan
- Integrate repository in all aspects of library operations (collection development, acquisitions, cataloging, reserves, reference, information literacy, liaison program)
- Build the ability for staff to be able to manage through uncertainty by training in change management using a multiplicity of methods
- Support staff through professional development opportunities
- Recognize that the type of staffing needed will vary depending on the software solution, whether commercial or open source, remotely hosted or supported locally. There are job descriptions and associated cost estimates available from several of the early adopter sites, such as MIT. ⁴

Lesson 6: Content submission

- Faculty have department or administrative staff who might be able to submit content
- Set expectation that new communities will launch in bi-annual releases to motivate continuing development
- Provide frequent training sessions to content submitters on entering metadata
- Authority control is a large issue in metadata submission. Try using dropdown menus for author names, department names, etc
- In addition to faculty, train graduate students in submitting content

Lesson 7: Publicity and marketing

- Ongoing communication throughout campus, timed to match academic calendar
- Continued contact with existing communities to provide support, training, move project along
- Use success stories, with quotes from early adopters, to market service to other faculty to recruit new content
- Work with Office of Grants and Contracts Administration to reach faculty that need to show funding agencies how their work will be distributed and preserved
- Invest time and money to get a community pilot started
- Address each discipline's needs using its own vocabulary, content examples, and grey literature
- Educate faculty in the topics of scholarly communication, digital preservation, etc and how an IR can help them
- Demonstrate to faculty the ease of submitting content, using the repository to find content
- Emphasize the persistent identifier for content, a key selling point for faculty
- Word-of-mouth publicity among faculty and other users of repository. Tap into this resource providing accurate information
- Success is not usually immediate

³ BARTON M.R. and WATERS M.M. *Creating an Institutional Repository*. <<http://dspace.org/implement/leadirs.pdf>>.

⁴ BARTON M.R. and WATERS M.M. *Creating an Institutional Repository*. <<http://dspace.org/implement/leadirs.pdf>>.

Why Libraries? Why Librarians?

Libraries have a strategic and tactical role to play in moving beyond their traditional custodial role to active engagement in the scholarly communication enterprise. Librarians traditionally have expertise in numerous areas that can be applied to IRs. Here are a few examples of roles that we can exploit in new ways.

- Library role as steward of scholarship
- Collection development expertise
- Liaison role with faculty
- Expertise on metadata and format management
- Commitment to long-term preservation
- Complementarities of repository and licensed digital materials
- Dissemination and access management

Exploration of New Roles and Partnerships

As we look at the transformation of digital publishing and its impact on the role of libraries and librarians in the 21st century, Kennan and Kautz have excellent model diagrams that depict the shift from traditional to electronic scholarly publishing, then provide a conceptual model of the possibilities for publishing enabled by IT ⁵

What are some of these new models?

- Open access journals are journals that are available at no cost to the end user, may or not be refereed, and typically are available in electronic format.
- Institutional digital repositories provide unified open access to and preservation of the electronic collections of works of members of the institution's community.

What does this mean for the transformation of the traditional roles that we just discussed? Can we do this?

Looking at the experiences of early IR adopters and the trends in scholarly communication, it's clear that libraries and librarians must consider that providing IR services is a critical component of their mission. By exploring some of the ways that we can lead the academy in these efforts, we begin to see how librarians' roles are effective in managing this change and how these roles will become critical to the success of the university over time.

- Organizing and maintaining digital content
- Support for faculty as information content contributors and end users
- Support faculty's open access digital publishing activities.
- Document preparation expertise (format control, archival standards, etc)
- Expertise in metadata tagging, authority control, content management to increase access to and use of data
- Document management (new uses for collection development expertise)

⁵ KENNAN M.A. and KAUZ K. *Scholarly Publishing and Open Access*. <<http://dlist.sir.arizona.edu/1867/>>.

- Critical faculty and administration outreach and education to best showcase the institution's research
- Work closely with faculty, administrators, and other stakeholders

Librarians will have an extraordinary visibility within the university as they take on these roles.

Persistent throughout this list of new roles for libraries and librarians is working directly with faculty in new ways. Past experience indicates that faculty are not familiar with how to manage their copyrights and how to retain their author rights. In their need to get published for promotion and tenure, they tend to sign current publisher agreements without reading them thoroughly. Many librarians have developed expertise in the use of Creative Commons licenses and scholar's copyright addenda as mentioned previously. They also need guidance in discovering journals and publishers with access friendly policies so another service that librarians can provide is to determine open access journals in the research areas of their faculty <<http://doaj.org>> and investigate current publisher policies using available tools such as the Sherpa/RoMEO website <<http://www.sherpa.ac.uk/romeo.php>> and publisher websites.

Faculty need to be encouraged to examine their own roles in the current publishing system be it as an author, editor, member of an editorial board, or as a peer reviewer. They often do not realize the negotiating power they bring to the table in these respective roles. Faculty can also be encouraged to prefer open access grant agencies when seeking funding (Wellcome Trust) <http://www.wellcome.ac.uk/doc_WTD002766.html> and to self-archive in their institution's digital repository and in disciplinary repositories such as arXiv.org <<http://arxiv.org/>> and RePEc <<http://repec.org/>> in the States.

Examples from Case Study

Let's take a look at some examples of the topics that we've been discussing this morning, starting with a look at the website we created to address several aspects of scholarly communication for a variety of audiences. <<http://guides.library.umass.edu/scholarlycommunication/>>.

Collaboration with the University Press began early in our pilot projects. The Press contributed several books by UMass Amherst authors to the digital repository and they are some of the most heavily downloaded materials. <<http://scholarworks.umass.edu/umpress/>>.

A faculty member who was publishing a monograph with a University Press had extensive primary materials and footnotes that they could not include in her publication. They suggested that she create a website. She remembered the presentation about ScholarWorks at her faculty department meeting and put this unique material into the IR. <http://scholarworks.umass.edu/french_translators/>.

A faculty member who is an editor of one of the open access journals on our IR, showcased it to his professional association. This won an award for UMass Amherst for supporting their scholarly communication endeavors. We

now support two more of their journals. After this success, the professional association decided to investigate the “events” functionality of the IR and is now in the process of migrating their entire conference process, from call for papers to final schedule and online proceedings, to the IR. <<http://scholarworks.umass.edu/jhte/>>.

Another early partner in our IR development is the Graduate School which began its partnership with the library by including PhD dissertations. We then explored the possibility of providing the software, rights management and format support for electronic masters theses as well so they’d have a comprehensive listing of all degrees. Additional functional allows for communities of ETDs to be embedded in departments’ or programs’ IR websites so their scholarship can be showcased to prospective faculty and graduate students:

<<http://scholarworks.umass.edu/etds/>>,

<<http://scholarworks.umass.edu/anthro/>>.

The University’s Office of Research, mentioned previously as another strategic partner in our IR development, has several examples of mandated deposits from locally awarded research grants:

<<http://scholarworks.umass.edu/cleanenergy/>>,

<<http://scholarworks.umass.edu/timbr/>>.

This partnership has also led to the creation of new interdisciplinary open access journals evolving from the mandated conference deposits.

Conference: <<http://scholarworks.umass.edu/violence/>>.

Journal: <<http://scholarworks.umass.edu/lov/>>.

New strategic partnerships have included the Office of Outreach and their community engagement activities. UMass Amherst is a land-grant institution in the US which means that a critical part of the university’s mission is to provide agricultural and technical education and promote higher education, especially to rural and underserved parts of the population.

<<http://scholarworks.umass.edu/engagement/>>.

Concluding Remarks

David Shulenburg, the closing keynote speaker at the SPARC Digital Repositories Meeting in November 2008, described it best. “A digital repository is the most effective way to make information free and the public wants and desperately needs to have it.” He suggested that the value of digital repositories is in “building interest in the scholarship of a specific faculty member as measured by citations”, “enhance the reputation of the institution with funding agencies,” “enable the institution to fully understand the breadth and depth of the work in which faculty are engaged” and “preserve scholarship that otherwise might be lost”⁶

He suggests several steps for moving forward that have been incorporated into several organizations’ strategic

plans including the Association of American Universities, ARL and the Coalition for Networked Information.

1. Make sure there is a digital repository available for deposit work of your university’s faculty.
2. Work with the president, provost and faculty to show real benefits of broadening distribution of scholarly product.
3. Initiate discussions with administration and faculty to modify current practices and intellectual property policy so university retains certain rights.
4. Support efforts to spread public access policies, such as those of NIH, to all funding agencies and foundations.
5. Work to educate campus units to convince them to support, not to oppose the best interest of members.
6. Work with departments and faculty to develop habits of depositing into digital repository.
7. While information in a digital repository and search engines may sell itself, brand the information in your repository as information your university created or made available to the public. Work with your university public relations unit to spread the word so the public looks first to your repository for reliable information.

“Digital repositories can make cultural contributions of scientific knowledge generated on campuses widely available and have profound effects,” Shulenburg says. “This is a case of doing good, while doing well.”

References

- Association of Research Libraries (ARL). *The Research Library’s Role in Digital Repository Services* [online]. Final report of the ARL Digital Repository Issues Task Force. January 2009 [cited 2009-05-18]. Available from: <<http://www.arl.org/bm~doc/repository-services-report.pdf>>.
- Association of Research Libraries (ARL). *Transformational Times* [online]. An Environmental Scan Prepared for the ARL Strategic Plan Review Task Force. Feb 2009 [cited 2009-05-18]. Available from: <<http://www.arl.org/bm~doc/transformational-times.pdf>>.
- Prepared to support ARL’s strategic planning initiative, the document identifies key trends affecting public policy issues, scholarly communication, and emerging roles for libraries in research, teaching, and learning.
- BAILEY, Charles W. *Institutional Repositories, Tout de Suite* [online]. 2008 [cited 2009-05-18]. Available from: <<http://www.digital-scholarship.org/ts/irtoutsuite.pdf>>
- BARTON, Mary R and WATERS, Margaret M. *Creating an Institutional Repository* [online]. LEADIRS Workbook. Boston: MIT Libraries, 2004-2005 [cited 2009-05-18]. Available from: <<http://dspace.org/implement/leadirs.pdf>>.
- Step-by-step guide on how to build an institutional repository. It contains worksheets, examples, planning guides, cost modeling, and references to other sites.
- Council on Library and Information Resources (CLIR). *No Brief Candle : Reconceiving Research Libraries for the 21st Century* [online]. Aug 2008. CLIR Publication no. 142 [cited 2009-05-18]. Available from: <<http://www.clir.org/pubs/reports/pub142/pub142.pdf>>. ISBN 978-1-932326-30-7.

⁶ SHULENBURGER D. Closing keynote. <http://www.arl.org/sparc/meetings/ir08/closing_keynote.shtml>.

Essays in this collection identify leading issues to be considered when imagining the roles and functions of the academic library in the 21st century.

CROW, Raym. *The Case for Institutional Repositories: A SPARC Position Paper* [online]. SPARC, 2002 [cited 2009-05-18]. Available from: <http://www.arl.org/sparc/bm~doc/ir_final_release_102.pdf>.

Describes IRs and focuses on two areas: role in the changing scholarly communication arena and in being a showcase of an institution's research and public value. Also explores IRs impact on major scholarly communication stakeholders.

CROW, Raym. *SPARC Institutional Repository Checklist & Resource Guide* [online]. 2002 [cited 2009-05-18]. Available from: <http://www.arl.org/sparc/bm~doc/ir_guide__checklist_v1.pdf>.

Raym Crow's practical guide offers an overview of the major issues which institutions and consortia need to address in implementing an institutional repository. Contains a more comprehensive bibliography.

FOSTER, Nancy Fried and GIBBONS, Susan. Understanding Faculty to Improve Content Recruitment for Institutional Repositories. *D-Lib Magazine* [online]. Jan 2005, vol. 11, no. 1 [cited 2009-05-18]. Available from: <<http://www.dlib.org/dlib/january05/foster/01foster.html>>. ISSN 1082-9873

FYFFE, Richard, LUDWIG, Deborah and WARNER, Beth Forrest. Digital Preservation: A campus-wide perspective. EDUCAUSE Center for Applied Research. *Research Bulletin* [online]. 2005, no. 18 [cited 2009-05-18]. Available from: <<http://www.educause.edu/LibraryDetailPage/666?ID=ERB0518>>.

FYFFE, Richard and WARNER, Beth Forrest. *Scholarly Communication in a Digital World* [online]. The Role of an Institutional Repository. University of Kansas. March 2003. <<http://kuscholarworks.ku.edu/dspace/handle/1808/126>>.

Richard Fyffe and Beth Forrest Warner describe the role that repositories can play in scholarly communication, and the benefits they offer to the system.

GIBBONS, Susan. Establishing an Institutional Repository. *Library Technology Reports* [online]. July - August 2004, 40:4 [cited 2009-05-18]. Available from: <<https://publications.techsource.ala.org/products/archive.pl?article=2538>>.

This is a „soup to nuts“ resource for anyone seeking more information on institutional repositories. She outlines the steps for establishing an institutional repository, its features, uses, available systems, costs, and much more.

JOHNSON, Richard K. Institutional Repositories: Partnering with Faculty to Enhance Scholarly Communication. *D-Lib Magazine* [online]. Nov. 2002, vol. 8, no. 11 [cited 2009-05-18]. Available from: <<http://www.dlib.org/dlib/november02/johnson/11johnson.html>>. ISSN 1082-9873.

Clarifies strategic roles an IR serves for colleges and universities, describes IR's role in scholarly communication, defines essential elements and explains what needs they serve for faculty and researchers.

KENNAN, Mary Anne and KAUZ, Karlheinz. Scholarly Publishing and Open Access: Searching for Understanding of an Emerging Phenomenon. In: *Proceedings ECIS 2007 — The 15th European Conference on Information Systems* [online]. St Gallen, Switzerland [cited 2009-05-18]. Available from: <<http://dlist.sir.arizona.edu/1867/>>

KORNBLUH, Mark Lawrence. Envisioning the Future of the University in the Digital Age: Thinking Big About Institutional Repositories. *CNI-JISC-SURF Conference*, 11 May 2005, Amsterdam, Netherlands.

LINDAHL, David and FOSTER, Nancy. Use a Shoehorn or Design a Better Shoe: Co-Design of a University Repository. *Participatory Design Conference* [online]. Toronto, ON, Canada, 29-31 July 2004 [cited

2009-05-18]. Available from: <<https://urresearch.rochester.edu/retrieve/3237/Participatory+Design+Conference+Paper+2004.07.31.pdf>>.

Explores new web services that faculty need in order to effectively utilize an IR.

LYNCH, Clifford A. Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age. *ARL: A Bimonthly Report* [online]. February 2003, no. 226 [cited 2009-05-18]. Available from: <<http://www.arl.org/resources/pubs/br/br226/br226ir.shtml>>.

Provides a comprehensive overview of institutional repositories, from defining what they are, to clarifying their strategic importance, to some cautions to consider.

MACKIE, Morag. Filling Institutional Repositories: Practical strategies from the DAEDALUS Project. *Ariadne* [online]. April 2004, no. 39 [cited 2009-05-18]. Available from: <<http://www.ariadne.ac.uk/issue39/mackie/>>.

MCCORD, Alan. *Institutional Repositories* [online] : Enhancing Teaching, Learning, and Research. EDUCAUSE Evolving Technologies Committee, 16 October 2003 [cited 2009-05-18]. Available from: <<http://www.educause.edu/ir/library/pdf/DEC0303.pdf>>.

RIEH, Soo Young, et al. Census of Institutional Repositories in the U.S.: A Comparison Across the Institutions at Different Stages of IR Development. *D-Lib Magazine* [online]. Nov 2007, vol. 13, no. 11/12 [cited 2009-05-18]. Available from: <<http://www.dlib.org/dlib/november07/rieh/11rieh.html>>. ISSN 1082-9873.

SHULENBURGER, David. *Closing keynote* [online]. SPARC Digital Repositories meeting, Baltimore, MD, Nov 2008 [cited 2009-05-18]. Available from: <http://www.arl.org/sparc/meetings/ir08/closing_keynote.shtml>.

Website resources

ALA / LITA. *Establishing an Institutional Repository* [online]. 14 January 2005, Boston, MA [cited 2009-05-18]. Available from: <<http://www.ala.org/ala/lita/litaevents/midwinter2005boston/IR.htm>>.

ARL / OLMS Webcast. *Institutional Repositories : Revealing Our Strengths* [online]. Cosponsored by SPARC & CARL, 10 June 2004 [cited 2009-05-18]. Available from: <http://www.library.arizona.edu/documents/dlist/IR_final.pdf>.

Budapest Open Access Initiative (BOAI) [online]. 2002 [cited 2009-05-18]. Available from: <<http://www.soros.org/openaccess/initiatives.shtml>>.

Provides links to important initiatives that support the open access movement which have occurred since the launch of the BOAI in February 2002.

Creative Commons [online]. [Cited 2009-05-18]. Available from: <<http://creativecommons.org/>>.

EDUCAUSE. Resource Center [online]. Section on IRs [cited 2009-05-18]. Available from: <<http://www.educause.edu/Resources/Browse/-InstitutionalRepositories/27900>>

IMS Global Learning Consortium [online]. Digital Repositories Specifications, 2003-2005 [cited 2009-05-18]. Available from: <<http://www.imsglobal.org/digitalrepositories/index.html>>

Provides recommendations for interoperation of the most common repository functions. Includes three sections: Core Functions information model, XML binding specifications, and best practices and implementation guide.

OpenDOAR (Directory of Open Access Repositories) [online]. [Cited 2009-05-18]. Available from: <<http://www.opendoar.org/>>.

SHERPA (Securing a Hybrid Environment for Research Preservation and Access) [online]. [Cited 2009-05-18]. Available from: <<http://www.sherpa.ac.uk/>>.

Includes the SHERPA / RoMEO Publishers Copyright listings, advice on creating a repository, advocating, submission guidelines, preservation issues, licenses and policy development in the UK. It focuses on open source software platforms EPrints and DSpace.

SPARC-IR Discussion List [online]. [Cited 2009-05-18]. Available from: <<https://mx2.arl.org/Lists/SPARC-IR/>>.

An online discussion list of institutional repositories. Participants can ask questions, share best practices and debate relevant issues.

SPARC Open Access Newsletter [online]. [Cited 2009-05-18]. Available from: <<http://www.earlham.edu/~peters/fos/>>.

A monthly bulletin with analysis of open-access developments put together by noted advocate Peter Suber.

SPARC Repository Resources [online]. [Cited 2009-05-18]. Available from: <<http://www.arl.org/sparc/repositories/>>.

SPARC / SPARC EUROPE [online]. *Institutional Repositories: The Next Stage*. 18-19 November 2004, Washington, DC [cited 2009-05-18]. Available from: <<http://www.arl.org/sparc/meetings/ir04/ir04.html>>

Topics: how to populate the IR, policy and legal issues (authors' agreements, licenses, copyright), business models, digital preservation, and various technical solutions.

The End of Institutional Repositories & the Beginning of Social Academic Research Service: an Enhanced Role for Libraries

Stuart Basefsky

smb6@cornell.edu

Cornell University, USA

Abstract: As more and more universities establish Institutional Repositories (IR), awareness is developing about the limitations of IRs in enhancing the academic research service. The concept of an IR needs to be expanded to include the integration of the processes that transform intellectual endeavor into a broadening array of academic and research support services which are fundamentally social. These include, but are not limited to sharing institutionally developed intellectual product (traditional IR); informing others of the availability of this product with defined purpose; collecting additional academically relevant materials in digital formats using IRs; disseminating timely information about what has been collected to researchers; creating an environment that encourages awareness and exchange of information; and more....

In brief, information gathering, dissemination, and discussion in the form of library service must become a crucial part of researchers' networks. An IR cannot and should not be viewed as a stand alone endeavor. It needs to be viewed and used as a research and communication tool in an environment that synergizes all elements of the research process. If an IR does not create discussions between librarians (information specialists) and researchers, its potential is lost both to the academy and the library. The library and its librarians must be interactive with researchers and the institution served.

Keywords: institutional repositories; digital publishing; new roles for librarians

The concept of the institutional repository (IR) is too narrowly focused to develop the value that universities should be extracting from its existence. Is it not possible for IRs to serve as full-fledged electronic libraries and thereby serve the greater purpose of collecting, disseminating, analyzing and exchanging useful digital information for academic purposes? Should not the IR be coupled with the full range of academic and research support services that new technologies permit? In an era of social networking, why is the university not moving quickly to develop what I call a social academic research service that can enhance the role of libraries, librarians, and IT specialists in the academic endeavor? It is the assertion and questions above that I hope to address in this presentation. Many of the points will be made by using the example of the Catherwood Library of the ILR School (School of Industrial & Labor Relations) at Cornell University.

The basic issues that I will cover are as follows:

1. How did IRs become so narrowly focused? – including a brief background on IR development.
2. How marketplace developments in IR technology allowed for a broader conception of use for a library that did not lose sight of its core business.
3. Using the new technology to reinvent academic and research support services.
4. Developing new directions and possibilities for support of the academic endeavor

How did IRs become so narrowly focused?

Too often institutions and individuals are enamored with the changes and opportunities that new technology offers. This was evident during the “dot.com revolution” of the 1990s.¹ This was a time of great technological progress and prospect. The hype which drove a stock market boom ended in a bust, however. It should be noted that this was the same time that the concept of the institutional repository (IR) in academia was born. Unfortunately, the clarity of vision that a “bust” forces upon investors was never experienced by academics who did not have to respond to bottom-line realities. The investments in innovation were not constrained by the overriding question of, “What value are we (the universities) to gain from this investment?” As a result, the personal interests and agendas of researchers in fields related to information technology were allowed to flourish often with great benefits for the world at large. But what benefits came to the processes of academic endeavor and inquiry? This question was rarely asked. After all, the academy is made up of brilliant individuals with personal research agendas that fit under the umbrellas of various disciplines and departments within the larger university and the world of academia in general. Whatever concerns there were about technological investments were guided more by

¹ WAHLGREN, E. *Legacies of the Dot-Com Revolution*. <http://www.businessweek.com/careers/content/mar2001/ca20010320_628.htm>.

individual interests rather than by institutional goals or the goals of the academy at large.

The development of the IR [see the BACKGROUND section below] reflects well the pattern described above. Indeed, the IR attracted researchers to develop robust technologies that would be interoperable. These researchers were, however, more interested in the technology questions than in the sustainability and value questions that should have been of interest to their institutions. They came up with competing systems that often required their own institutions to invest heavily in their areas of inquiry with the promise that the IR would serve the interests of their fellow faculty members throughout the university. This was an assertion that was true but short sighted. The IR was developed and sold to universities as a way to extend the reach of the intellectual products that were being developed at the university as a whole. In brief, this rationalized the role of the IR for a campus but with a focus on the personal faculty agendas of the IR developers and the needs of their fellow faculty to self-promote and by extension promote the usefulness of the universities they all serve. This appeared to be a “win-win” situation. In many respects it was. The larger question should have been, “Is that all the value that we can extract from an IR?” Apparently this was enough for Harvard.² However, from my perspective and those of my colleagues at the Catherwood Library of the ILR School (School of Industrial & Labor Relations) at Cornell University, this is not nearly enough.

Background on IR development

The institutional repository (IR) concept was born out of competition for who was going to be responsible for dissemination of an institution’s intellectual product via the Internet. Individual institutions felt that the competitor in this instance was the e-print archive movement controlled by several well-funded or subscription-based groups, associations, and institutions serving a variety of disciplines. The key word that was used at the time was *interoperability*. This meant that a technical mechanism should be created and “designed to facilitate the discovery of content stored in distributed e-print archives” using an open archive architecture.³ The issues of concern were mainly technology, access (which should be free), and control of intellectual product. The IR was perceived to be an end unto itself. Every institution or a consortium of institutions would enter their e-prints into an IR and share useful academic and research products to enhance and stimulate study. The details of this process were much more complex than this description reveals. Copyright issues, institutional branding, peer review, faculty compliance and other challenges made the implementation difficult and costly.

² GUTERMAN L. *Harvard Faculty Adopts Open-Access Requirement*. <<http://chronicle.com/article/Harvard-Faculty-Adopts/40447>>.

³ Van de SOMPEL H. and LAGOZE C. *The Santa Fe Convention of the Open Archives Initiative*. <<http://www.dlib.org/dlib/february00/vandesompel-oai/02vandesompel-oai.html>>.

Libraries were perceived to be appropriate locations for these IRs since “information management” and the application of metadata were considered instrumental for successful implementation. Libraries welcomed this attention since they were fearful of being marginalized in light of web developments and the use of the Internet for information gathering. The IR would help the library maintain an important role in academic life in this time of disruptive technological change.

Unfortunately the focus on technology made most academic institutions and their libraries lose sight of the fundamental roles and interplay of universities, faculty, and librarians in the pursuit of knowledge. In addition, all three are now dependent on information technologists to support their functions. Moreover, information technologists are encouraged to offer innovative ideas for teaching, research and study. They are now a fourth party in the academic endeavor – a new development for the academy.

How marketplace developments in IR technology allowed for a broader conception of use for a library that did not lose sight of its core business.

While the focus of IRs remained narrow during the “dot.com revolution”, some concerns were developing about costs and sustainability. A clear example was the development of the partnership between the California Digital Library (CDL) and Berkeley Electronic Press (Bepress).⁴ The ensuing result of this partnership was the creation of the Digital Commons which started with the use of Bepress’s institutional repository software in 2002 for the California Digital Library’s eScholarship Repository.⁵ The cost issue was the main driver of this development. Sharing a single platform that would be used in common by several campuses would reduce costs. A business model was developed that reduced costs for each individual campus while standardizing several services. Nevertheless, a broader vision of the uses of an IR was not developing although some experimentation by individual schools and libraries was going on mainly in the area of capturing non-copyrighted materials produced primarily by U.S. government agencies and placing them in a few IRs.

When the Bepress IR software was being developed in 2002, the Catherwood Library of the ILR School at Cornell University (henceforth Catherwood Library) was looking for a way to serve better its faculty, researchers, students

⁴ Bepress (Berkeley Electronic Press). *California Digital Library and Berkeley Electronic Press Announce Partnership for Scholarly Communication Initiatives*. <<http://www.bepress.com/press100401.html>>.

⁵ Bepress. *Customers Using the Digital Commons Platform*. <<http://www.bepress.com/ir/customers.html>>.

and world-wide audience.⁶ The concern was to capture not only the intellectual product of the school, but to provide the full range of services that the library envisioned for a modern electronic library service (exclusive of the library online catalog). This included capturing digitally born information, digitizing useful print materials for broader dissemination, creating electronic collections of workplace related material, providing awareness of recently gathered information and more. In brief, the question that was raised was essentially, “Can we use an IR type technology to be an electronic bookshelf?” This library, through my own efforts, had been struggling with this problem since 1994 when it started placing the U.S. Glass Ceiling Commission materials on its web site.⁷ Awareness of Bepress success was already evident at the Cornell Law Library which was using the Bepress run NELLCO Legal Scholarship Repository for its own purposes.⁸ That awareness led me to suggest that we use Bepress (which would be bought out by Proquest – later to revert back to Bepress) and its IR software for this purpose. However, it was not until 2004 that Catherwood Library would commit to what was then the Proquest/Bepress Digital Commons.⁹

While the above paragraph provides a chronology of actions, it was the culture of the Catherwood Library that would eventually lead to the success of the IR and expand its usefulness into what I call a “social academic research service” (to be explained in the following paragraphs). In fact, the foundations of a successful IR had been built years earlier.

Since 1946, when the Catherwood Library was established¹⁰, it was viewed as an integral part of the research and outreach endeavor as the provider of information in support of those functions. The library was in partnership with the school and its faculty to fulfill the broad mission that it was assigned by the State of New York. In order to be effective, the library went beyond the normal standards of providing good service. In its early years, it created a thesaurus to compensate for the inadequacy of Library of Congress subject headings as they related to industrial and labor relations and workplace issues in general so that added entries could be made in the catalog that would make material easier to find. In addition, articles from journals not adequately covered by journal indexing services were cataloged for the use of faculty, researchers, students, and the public. Strong linkages between faculty and those responsible for collection

development and reference services were also developed. Faculty often discussed their issues and concerns with the librarians. In fact, faculty often went to lunch with the librarians including them in their network of academic endeavor. These are but a few examples of making extraordinary efforts to distinguish not only the library as an excellent service provider but the school that it served as a pre-eminent institution in its field. Hence, a culture of cooperation was long present. In brief, the Catherwood Library was part of the social fabric of the institution it served.

As the technology was changing the world of libraries and the habits of information seekers, this culture allowed the Catherwood Library to embrace changes to further enhance the special “social role” it had developed with its primary clientele. In fact, numerous discussions within the library had taken place about how we could maintain our already influential position within our institution. As a consequence, a variety of experiments were undertaken. I wrote about many of these and presented our library’s developing view (my view in particular) of the new roles that a library could undertake on behalf of its institution.¹¹ Moreover, having recognized the trends and developments that were changing the way libraries would have to operate in order to maintain their effectiveness and utility, I set in motion a conceptual framework of how to operate in this new environment outlined in an article entitled, “*The Library as an Agent of Change: Pushing the Client Institution Forward.*”¹²

As a result, when the “dot.com revolution” was ending the monopoly on information long held by libraries, the Catherwood Library was prepared to incorporate new technologies to enhance its role and function within the school that it served. In fact my colleagues at the time, Constance Finlay and John Goddard, created a website in 1994 not only for the library, but for the school itself. This ultimately became the ILR School’s web site of which the Catherwood Library is now only a part.¹³ From the outset of creating a web site, we began collecting not only our faculty’s intellectual product for viewing and dissemination, but non-copyrighted material that was relevant to our field.¹⁴ Alas, these early efforts came upon a technical ceiling – the size and cost of digital storage. This forced us to be on the lookout for new and developing technologies to better serve our school.

It was at this point, that we had a long wait before IR developments became robust enough and reasonably priced for us to consider using them for our long hoped

⁶ ILR School [Online]. About Catherwood Library. <<http://www.ilr.cornell.edu/library/about/>>.

⁷ GEDDES D. Library-government partnership will aid access. <<http://www.news.cornell.edu/Chronicle/96/1.18.96/ILRarchive.html>>.

⁸ NELLCO (New England Law Library Consortium) Legal Scholarship Repository. <<http://lsr.nellco.org/>>.

⁹ Bepress. Berkeley Electronic Press and ProQuest Announce New Institutional Repository Service. <<http://www.bepress.com/press062504.html>>.

¹⁰ Martin P. Catherwood Library. <http://en.wikipedia.org/wiki/Catherwood_Library>.

¹¹ BASEFSKY S. *The Influential Library and Librarian: Practice and Research Agenda for the 21st Century*. <<https://confluence.cornell.edu/display/prodev/PD+Week+Archive#PDWeekArchive-PDWeek1999>>.

¹² BASEFSKY S. *The library as an agent of change*. <<http://digitalcommons.ilr.cornell.edu/articles/189/>>.

¹³ ILR School (School of Industrial and Labor Relations) web site. <<http://www.ilr.cornell.edu/>>.

¹⁴ See ILR School web site. Nov. 6, 1996 through the Internet Archive (Wayback Machine). <<http://web.archive.org/web/19961106150649/http://www.ilr.cornell.edu/>>.

for electronic library. However, once this materialized, our library's culture and its long maintained focus on socially integrating ourselves into the academic endeavors of our faculty made the introduction of an IR a relatively simple transition compared to other libraries and schools as explained by Suzanne Cohen (current colleague) and Deborah Schmidle (former colleague now working at Cornell's central library).¹⁵

The Catherwood Library made the commitment to use the Digital Commons in 2004.¹⁶ Our usage statistics are extraordinary compared to other libraries and schools as a result of holding onto our social relationships with the school that we serve.¹⁷ Faculty give us their material willingly so much so that we are on occasion overwhelmed with their insistence on incorporating their materials both current and historical into our IR. However, this IR is not viewed as an IR by our school. This IR has become a communications vehicle useful to our scholars on numerous levels. It is not used simply for self or institutional aggrandizement. It is an interdisciplinary research and outreach tool focused on workplace issues.

Is the Catherwood Library an exception? Are there other libraries similarly situated with this kind of potential?

Skeptics might think that the social networking that exists at the ILR School is not applicable to large academic research libraries. However, they are incorrect. The model for creating socially integrated libraries in research institutions is the foundation of what exists (or should exist) in professional schools in the USA such as law schools or medical schools. Other specialized libraries in engineering, business, hotel, agriculture, and more can point to similar dynamics. The problem, from my perspective, is that large academic research libraries have neglected to think of themselves as the "special library" of their institution – the university. In fact, the Catherwood

Library has tried to set an example for how large institutions can leverage their expertise to train not just faculty and students, but administrators and their staff so that they might operate more efficiently and become more competitive. Is it not possible to harness the expertise of librarians to target the special priorities of a university or key faculty research endeavors? The Catherwood Library made specific efforts to train the human resource personnel at Cornell University to enhance the performance and productivity of that staff. This is something that I have written about¹⁸ and others have noted.¹⁹ Certainly one would think that library institutional outreach could be improved with an eye toward developing the social academic research services outlined above.

If this is not taking place, perhaps there are other factors that impede these developments. For example, the majority of large academic research libraries are run by persons primarily with a technical services background. Their personal experiences do not include extensive contacts with faculty and researchers. They may be more comfortable using technology in providing access and other passive services rather than exerting pro-active services which engage users in intellectual give and take. Passive services and pro-active services need to be properly balanced. One should not dominate the other. If these leaders are not comfortable reaching out, they should simply ask for assistance in doing so.

The key difference in providing service in new technological environments is that clientele should not be expected to come to the library. They may choose to and this should be encouraged. The library needs to use technology effectively in reaching out to users. In the academy, this means bringing innovation to our thinking.

Using the new technology to reinvent academic and research support services.

Too many libraries take the attitude that "if they build it, they [users] will come." The questions that should be asked, however, are: Build what? For whom? Why? In building IRs, the evidence is clear that their mere existence does not translate into use. Hence the necessity to come up with Harvard-like mandates to force compliance of faculty. The social, academic foundation for cooperation and active participation in IR efforts was overlooked. However, a lesson can be learned from these failings: choose technologies that fit the culture of your organization so that participation by faculty, researchers, students, administrators, and staff will be more likely to take place. If

¹⁵ COHEN S. & SCHMIDLE D. *Creating a multipurpose digital institutional repository. and a shorter piece at NEWHART M. & COHEN S. DigitalCommons@ILR: Beyond the Traditional Institutional Repository.*
<<http://digitalcommons.ilr.cornell.edu/articles/118>>

¹⁶ Mary Newhart (former colleague) was instrumental in managing the new technology and articulating why the Digital Commons was our IR of choice. See:

NEWHART M. *ILR Review Presentation - Charleston Conference. (Presented by Mary Newhart, Web and Digital Projects Manager, Catherwood Library, School of Industrial and Labor Relations Cornell University. As part of the panel on The Confluence of Open Access and Licensed Content, 2005 Charleston Conference: Issues in Book and Serial Acquisition, November 2-5, 2005.)*,
<http://works.bepress.com/mary_newhart/1/>.

and

NEWHART M. *DigitalCommons@ILR Collection Development Policy.* Nov. 2005. <http://digitalcommons.ilr.cornell.edu/dc_information/4/>.

¹⁷ *DigitalCommons@ILR* <<http://digitalcommons.ilr.cornell.edu/>>. 8,796 papers (as of 5 May 2009) 1,527,592 full-text downloads (as of 5 May 2009) 643,366 downloads in the past year (as of 5 May 2009)

¹⁸ BASEFSKY S. *The other client : Information training for administrators pays dividends for the library.*

¹⁹ HURST L. *Academic Library as Special Library: A Model of Library Orientation for Campus Employees.*

and

HURST L. *The Special Library on Campus: A Model for Library Orientations Aimed at Academic Administration, Faculty, and Support Staff.*

a library does not know or perceive what its role (or perhaps developing role) is in helping the institution it serves, it can become an ineffective follower of the crowd or at worst a counter-productive cost center. So how does one choose proper technologies?

Most libraries try to benchmark what others in their field are doing. This can be informative. However it may not be creative or useful for the library seeking the information. In addition, it may not translate into providing effective leadership. “Following benchmarks and best practices that are not a good fit for your [university] or its culture can be counterproductive. The most effective way of using benchmarks and best practices is as a creative mechanism for raising questions about your own [situation]. Following what others do is rarely a form of good leadership.”²⁰

Assuming that a library does know its environment well, what can it do with the endless variety of communications technologies that can be incorporated into IRs or otherwise support information needs? The choices are enormous: discussion boards, blogs, rss feeds, interactive data sets, web pages, chat rooms, podcasts, and many more. However, a good reference librarian always tries to discern the “needs” of the user before seeking the appropriate answer. A good library should do the same. In my opinion, this is where one should start. There is a general tendency to focus on “user behavior” — the technologies currently fashionable among young adults and researchers. This can be instructive and it should lead to some experimentation. But the fundamental three questions remain: Build what? For whom? Why?

While I cannot speak for others, the Catherwood Library has made the following reasoned choices in reinventing the academic and research support services. Many of these choices come from accepting the current and future reality that Internet-based information, as made available via Google and other search engines, provides significant primary research material far exceeding anything that can be collected by a university or school. Given the temporary nature of the Internet, however, it is necessary to collect, disseminate, and retain key information necessary to a field or discipline for future users. Efforts to compete with Google or other search engines in creating an academic search portal should not be undertaken. However, an academic research portal that makes it easy to find and access the endless variety of formats that a research library must collect and retain whether purchased or acquired freely is a necessity. As a consequence, the Catherwood Library has become an ambidextrous library — providing both traditional and innovative services.²¹

Push E-Mail Services

A concerted effort to monitor, capture, and disseminate key information released daily via the Internet is made. At the ILR School this is called the IWS Documented News Service.²² Its intended purpose is “to solve a key public policy issue: how to disseminate workplace-related studies, reports, and data issued by government bodies (international, federal, state, and local) so that they reach interested faculty, researchers, practitioners, students, journalists, and the general public while the content is still current [and much more]”²³

Reasons:

Speed — Historically in the print world libraries would collect information as quickly as they could and incorporate it into a catalog. It was widely understood that academic inquiry on the subject matter would probably be delayed from a few months to a couple of years. For digitally born information available via the Internet, academic inquiry begins within a few minutes and may be delayed as much as a few days or weeks. Hence, it is necessary to collect and disseminate immediately. However, collection is at least a two step process. In this instance, the first collection takes place by an information specialist (Stuart Basefsky) who then disseminates the information via e-mail. One of the e-mail addresses places the e-mail directly and automatically into at least one of three **blogs**.²⁴

Library Synergies — Once initially identified above, the information is also e-mailed to the collection development librarian who may decide to collect a print equivalent, if available, or seek permission to incorporate the material into the IR for long-term access. The reference librarians also receive the e-mail which can help them keep current as well and permit them to answer inquiries with more current information. The reference librarians keep copies of each e-mail in a **searchable mailbox**. They use this as a **ready-reference database**.

Research Synergies — Many faculty do the same as the reference librarians. The service provides them with the possibility to keep only those items that interest them in a ready-reference database of their own making. Faculty share this information with their colleagues world-wide.

Note: This push e-mail service started in 1994, when I realized that with the Internet fewer people were going to come to the reference desk. I decided that if they were not going to come with questions, I was going to go to them with information that I knew from experience was useful. In other words, I was going to create my own questions and answers on their behalf and share it with them by

²⁰ BASEFSKY S. *Leadership & The Role of Information: Making The Creatively Informed Questioner*. <<http://www.ilrx.com/features/leadershipandroleofinformation.htm>>.

²¹ *InnovationPoint. The Ambidextrous Organization: A model for driving sustainable innovation within a mature business*. <<http://www.innovation-point.com/Ambidextrous.htm>>.

²² *IWS Documented News Service*. <<http://www.ilr.cornell.edu/iws/news-bureau/index.html>>

²³ *IWS Documented News Service. FAQ*. <<http://www.ilr.cornell.edu/iws/news-bureau/faq.html>>

²⁴ *Daily Postings* <<http://iwsdocumentednewsdaily.blogspot.com/>>. *International* <<http://iwsdninternational.blogspot.com/>>. *Weekly Bulletin* <<http://iwsdocumentednews.blogspot.com/>>.

creating a pro-active, continually updated bibliography (webliography). It was shocking at first. One faculty member said, "I knew that I needed to update a chapter of my book. Thanks to you, it arrived at my desktop unsolicited." The side effect was to train the faculty on how useful the Internet could be to their research.

Leveraging the Information into Academic Partnerships — Over time the news service picked up a reputation for disseminating the intellectual product of key research institutes and government offices. In exchange for bringing global attention to their academic and/or research contributions on an ongoing basis, these organizations were willing to take student interns and faculty exchanges. In fact, the news service is written into several Memoranda of Understanding (MOUs).²⁵ This kind of news service is one way to create institutional relationships. Too often university partnerships are dependent on individual faculty relations. The faculty are sometimes more concerned about their personal agenda than the institutional concerns. The library, through these kinds of services, can turn individual relations into institutional relations and sustain them.

Web-based Guides

These are quite common and need little explanation. However, the web guides produced by the Catherwood Library are quite extensive and worth looking at. They can be found at – <http://www.ilr.cornell.edu/library/research/>.

Reasons:

Select Key Resources by Topic — Members of the university community need both proprietary information as well as free information. These guides can combine both and refer to information that will never get cataloged by the library.

Highlight Key Resources for Classes — Working with faculty, special guides to help with class projects and assignments are created regularly.

Blog Use

Typically blog postings are intended to be of interest to patrons and feature new resources and services. At the Catherwood Library this kind of blog can be found at -

Catherwood Library @ ILR

<http://blogs.cornell.edu/catherwood/>.

Reasons:

Friendly Communications Tool — Blogs do not have to be bland. Creativity can be used to informally communicate with faculty and students. If desired, comments can be permitted (these would need monitoring, however).

RSS Feed Possible — Most blogs can be made to provide RSS access. Truly interested faculty and students can keep themselves current by using this feature.

IR Alerts

Information residing in the IR can be organized by category. These categories or even keywords can be set to push out alerts of newly added information of interest to researchers and others. At the Catherwood Library, the **DigitalCommons@ILR** serves as our IR (and more). The alert service is found at – <http://digitalcommons.ilr.cornell.edu/notification.html>.

Summations of News Stories for Current Awareness

Free alert services, providing abstracts and links to discipline-related news stories covered in the major media can be produced and maintained by students if they are properly trained. These can be configured so that a database archive can be searched of previously gathered stories. At the Catherwood Library this service is called **Workplace Issues Today (WIT)** found at – <http://www.ilr.cornell.edu/library/research/worldofWorkNews/wit/>.

Capturing [and Archiving] Key Web Sites for Research

Web sites are now the “new books” in that they bind the intellectual product of entire organizations into one identifiable element. These “archived web sites” can be placed in an IR and retained for research purposes long into the future. The Catherwood Library has now archived the entire International Labour Organization (ILO) web site as it existed on 31 January 2006. This can be viewed at – <http://ecommons2.library.cornell.edu/ilo/>. [This is still a work in progress].

Filtering and Highlighting Important Information

Implied by some of the examples above is the explicit need to provide filtering and highlighting of important information useful to researchers. As Sasha Skenderija pointed out at a previous CASLIN conference, "The urgent challenge to all information professionals is no longer availability and direct provision of resources, but rather the filtering and highlighting of the ubiquitous Infosphere."²⁶ Whether this filtering is done by using new technologies or by direct human intervention, it is a necessity. At the Catherwood Library, the reference and collection development librarians monitor the Internet for current information and couple this with technology that also grabs important reports from key information centers as they are placed on web pages – for example, using

²⁵ See: *Global Partners* <http://www.ilr.cornell.edu/international/servicelearning/globalPartners/>.

²⁶ SKENDERIJA S. *Law Library 2.0: New Roles for Law Librarians in the Information Overload Era.* <http://lsr.nellco.org/cornell/clsops/papers/36/>.

<http://www.watchthatpage.com/>>. Not enough of this is being done, although it has been advocated by many. It is a fundamental issue for managing libraries in the future.²⁷

MORE....?

There are now too many possibilities available. Using them wisely becomes an issue. One must also keep in mind the adage, “You better know what you want, because you might get it.” How does a library sustain a service that it creates? As we add and subtract some of these services at Catherwood, likely to appear or disappear on our Outreach pages at <http://www.ilr.cornell.edu/library/outreach/>>, concerns for efficient use of time and resources will be mitigated by technological developments it is hoped.

At the Catherwood Library emphasis is placed on maintaining intellectual relationships with faculty and students. Sometimes this is done via technology as listed above or by teaching classes or sessions of classes. The goal remains to be engaged with faculty and students in the academic endeavor. Whatever it is that we build, we know that it has to be viewed as being built *together* with all members of our community and for all members of the community. Building an IR, for example, without social academic underpinnings will not necessarily fail. It is likely to be less useful than anticipated, however. Too little value will be extracted relative to the cost and human efforts that go into it.

Developing new directions and possibilities for support of the academic endeavor.

This brings up the notion of where libraries are heading. Clearly they will be marginalized if they do not create their own futures and define the roles that they wish to play. It is now too easy for university administrators to view libraries as cost centers rather than value centers.²⁸ To my way of thinking, the only choice is to go back to the future.

Libraries were successful when they monopolized the print world. Unfortunately monopolies never have to explain themselves. Libraries, because of their monopoly, never had to articulate their roles in the academy even to themselves. They lost sight of the essential role that they played. I call this role the “social academic research service.”

The easiest way to explain a “social academic research service” is to think of it as blending the library and its services into the networks that faculty and researchers create for themselves in order to progress and succeed in their fields of inquiry and research. Libraries must become part and parcel of the organizations they serve. They must

be so intertwined as to be viewed as part of the whole that makes up a university. In brief, libraries need to learn what they have been trying to teach their users for millennia – quality and timely information is as important to an academic’s network as their colleagues and others with whom they socialize.²⁹

There is a problem in perception of libraries that comes from the fact that the university thinks of the library as a “thing” or a “place” rather than a set of services run by skilled and learned individuals. Libraries are, in fact, less important than the librarians and information technologists who through their learning and skills bring collected and organized and even chaotic information into play. The abilities that enable these specialists to “enhance serendipity” (my definition of librarianship) are what make these professionals true academics worthy of academic standing. To make it possible for faculty and researchers to discover useful information that they do not initially seek is the miracle work of librarians (and now information technologists). This has always been true.

The challenge, as I see it, is to keep librarians from undermining themselves. Libraries cannot afford to hide behind technology by creating passive services that emphasize access over real contact with real researchers. Technology permits the kind of contacts that create and enhance social networks of which the librarian and information technologist must be a part. However, creative contacts must include working with real people in clever ways. For example, in order to attract and retain top faculty some librarians and/or information technologists could be assigned to these targeted individuals and even written into their employment contracts as their “personal information trainers.”³⁰ Rather than being viewed as part of the infrastructure, librarians and IT personnel could become an employee benefit.

In any event, new technologies such as those designed to create IRs can be utilized in far more creative ways to enhance the research endeavor. I hope, therefore, to see the end of institutional repositories as we know them and the beginnings of a social academic research service. If done properly, there will be an enhanced role for libraries and a bright future for them as well.

Note: There is an irony in that by fostering a “social academic research service,” libraries will be making IRs more useful and attractive inadvertently as seen in the Catherwood Library example. As a result, the universities will get greater value out of their investment. Libraries will then be seen as agents of change, pushing their institutions forward.³¹

²⁷ GERMAIN, C.M. *Legal Information Management in a Global and Digital Age: Revolution and Tradition*. <http://www.llrx.com/features/legalinformationmanagement.htm>.

²⁸ BASEFSKY S. *Mis-information at the heart of the university: Why administrators should take libraries more seriously*. <http://digitalcommons.ilr.cornell.edu/articles/87/>.

²⁹ BASEFSKY S. *Proactive Leadership & The Role of Information: Identifying Strategic Networks of Information*. <http://www.llrx.com/features/proactiveleadership.htm>.

³⁰ BASEFSKY S. *The personal information trainer*. <http://digitalcommons.ilr.cornell.edu/articles/126/>.

³¹ BASEFSKY S. *The library as an agent of change: Pushing the client institution forward*. <http://digitalcommons.ilr.cornell.edu/articles/189/>.

References

- BASEFSKY, Stuart. *Leadership & The Role of Information: Making The Creatively Informed Questioner* [online]. 29 Oct 2008 [cited 2009-05-15]. Available from LLRX: <<http://www.llrx.com/features/leadershipandroleofinformation.htm>>.
- BASEFSKY, Stuart. Mis-information at the heart of the university: Why administrators should take libraries more seriously. *Information Outlook* [online]. Aug 2006, vol. 10, no. 8 [cited 2009-05-15]. p. 15-19. Free access from Cornell University ILR: <<http://digitalcommons.ilr.cornell.edu/articles/87/>>.
- BASEFSKY, Stuart. *Proactive Leadership & The Role of Information: Identifying Strategic Networks of Information* [online]. 12 Apr 2009 [cited 2009-05-15]. Available from LLRX: <<http://www.llrx.com/features/proactiveleadership.htm>>.
- BASEFSKY, Stuart. The Influential Library and Librarian: Practice and Research Agenda for the 21st Century. *Cornell University Library Professional Development Week* [online]. Inaugural Year, 18-21 May 1999 [cited 2009-05-15]. Conference site available from: <<https://confluence.cornell.edu/display/prodev/PD+Week+Archive#PDWeekArchive-PDWeek1999>>. Introduction: <https://confluence.cornell.edu/download/attachments/39878673/basefsky_99_intro.doc?version=1>. Presentation: <https://confluence.cornell.edu/download/attachments/39878673/basefsky_99.doc?version=1>. Handout entitled *Prescriptions for libraries seeking influential positions 12/98* <https://confluence.cornell.edu/download/attachments/39878673/basefsky_99_3.doc?version=1>.
- BASEFSKY, Stuart. The library as an agent of change: Pushing the client institution forward. *Information Outlook* [online]. Aug 1999, vol. 3, no. 8 [cited 2009-05-15] p. 37-40. Available after registration. Free access from Cornell University ILR: <<http://digitalcommons.ilr.cornell.edu/articles/189/>>.
- BASEFSKY, Stuart. The other client : Information training for administrators pays dividends for the library, *College & Research Libraries News*, Feb 2000, vol. 61, no. 2, p. 100-101.
- BASEFSKY, Stuart. The personal information trainer. *Information Outlook* [online]. Nov 2007, vol. 11, no. 11 [cited 2009-05-15], p. 10-17. Free access from Cornell University ILR: <<http://digitalcommons.ilr.cornell.edu/articles/126>>.
- COHEN Suzanne & SCHMIDLE, Deborah. Creating a multipurpose digital institutional repository. *OCLC Systems & Services*, 2007. Vol. 23(3), p. 287-296. Available online from Cornell University, School of Industrial and Labor Relations site: <<http://digitalcommons.ilr.cornell.edu/articles/118>> and a shorter piece at: *NEWHART Mary & COHEN Suzanne*. DigitalCommons@ILR: Beyond the Traditional Institutional Repository. *InsideCUL* [online]. May 2006 [cited 2009-05-15]. Available from: <<http://www.library.cornell.edu/insidecul/200605/digitalcommons.html>>.
- GEDDES, Darryl. Library-government partnership will aid access. *Cornell Chronicle* [online]. Jan. 18, 1996, vol. 27, no. 17. [Cited 2009-05-15]. Available from: <<http://www.news.cornell.edu/Chronicle/96/1.18.96/ILRarchive.html>>. ISSN 0747-4628.
- GERMAIN, Claire M. *Legal Information Management in a Global and Digital Age: Revolution and Tradition* [online]. Aug 27, 2007 [cited 2009-05-15]. Available from LLRX: <<http://www.llrx.com/features/legalinformationmanagement.htm>>.
- GUTERMAN Lila. Harvard Faculty Adopts Open-Access Requirement. *The Chronicle of Higher Education. News Blog* [online]. Feb 12, 2008 [cited 2009-05-15]. Available from: <<http://chronicle.com/article/Harvard-Faculty-Adopts/40447>>.
- HURST, Leslie. Academic Library as Special Library: A Model of Library Orientation for Campus Employees. *College and Undergraduate Libraries*. 2003, vol. 10, no. 1, p. 75-88.
- HURST, Leslie. The Special Library on Campus: A Model for Library Orientations Aimed at Academic Administration, Faculty, and Support Staff. *The Journal of Academic Librarianship*. Jul 2003, vol. 29, no. 40, p. 231-236.
- MARTIN P. Catherwood Library. In *Wikipedia. The free encyclopedia* [online]. (Redirected from Catherwood Library). [Cited 2009-05-15]. Available from: <http://en.wikipedia.org/wiki/Catherwood_Library>.
- NEWHART Mary. *DigitalCommons@ILR Collection Development Policy*. [online]. November 2005 [cited 2009-05-15]. Available from <http://digitalcommons.ilr.cornell.edu/dc_information/4/>.
- NEWHART Mary. *ILR Review Presentation – Charleston Conference*. [PowerPoint Presentation] Nov. 2005 [cited 2009-05-15]. Available from: <http://works.bepress.com/mary_newhart/1/>.
- SKENDERIJA, Sasha. *Law Library 2.0: New Roles for Law Librarians in the Information Overload Era*. Presented at the CASLIN conference, June 2007 [cited 2009-05-15]. Available from: <<http://lsl.nellco.org/cornell/clshops/papers/36/>>.
- SOMPTEL, Herbert van de and LAGOZE, Carl. The Santa Fe Convention of the Open Archives Initiative. *D-Lib Magazine* [online]. Feb. 2000, vol. 6, no. 2 [cited 2009-05-15]. Available from: <<http://www.dlib.org/dlib/february00/vandesompel-oai/02vandesompel-oai.html>>. ISSN 1082-9873.
- WAHLGREN, Eric. Legacies of the Dot-Com Revolution. *BusinessWeek. Employment Trends* [online]. 20 March 2001 [cited 2009-05-15]. Available from: <http://www.businessweek.com/careers/content/mar2001/ca20010320_628.htm>.

Website resources

- Bepress. Berkeley Electronic Press. [online]. *Customers Using the Digital Commons Platform*. [cited 2009-04-29]. Available from: <<http://www.bepress.com/ir/customers.html>>.
- Bepress. Berkeley Electronic Press. [online]. *Berkeley Electronic Press and ProQuest Announce New Institutional Repository Service*. Press Release 25 June 2004 [cited 2009-05-15]. Available from: <<http://www.bepress.com/press062504.html>>.
- Bepress. Berkeley Electronic Press. [online]. *California Digital Library and Berkeley Electronic Press Announce Partnership for Scholarly Communication Initiatives*. Press Release 4 Oct 2001 [cited 2009-05-15]. Available from: <<http://www.bepress.com/press100401.html>>.
- DigitalCommons@ILR [online]. [Cited 2009-05-15]. Available from: <<http://digitalcommons.ilr.cornell.edu/>>.
- ILR School (School of Industrial and Labor Relations) [online]. [Cited 2009-05-15]. Available from: <<http://www.ilr.cornell.edu/>>.
- ILR School [online]. *About Catherwood Library*. [Cited 2009-05-15]. Available from: <<http://www.ilr.cornell.edu/library/about/>>.
- ILR School web site [online]. 6 Nov 1996 [cited 2009-05-15]. Available through the Internet Archive (Wayback Machine) from: <<http://web.archive.org/web/19961106150649/http://www.ilr.cornell.edu/>>.
- ILR School [online]. Global Partners. [Cited 2009-05-15]. Available from: <<http://www.ilr.cornell.edu/international/servicelearning/globalPartners/>>.
- InnovationPoint. *The Ambidextrous Organization: A model for driving sustainable innovation within a mature business*. [Cited 2009-05-15]. Available from: <<http://www.innovation-point.com/Ambidextrous.htm>>.
- IWS Documented News Service [online]. [Cited 2009-05-15]. Available from: <<http://www.ilr.cornell.edu/iws/news-bureau/index.html>>.
- NELLCO. (New England Law Library Consortium). Legal Scholarship Repository [online]. [Cited 2009-05-15]. Available from: <<http://lsl.nellco.org/>>.

A Short Overview of International Trends in Open Access

Leo Waaijers

leowaa@xs4all.nl

Open Access consultant, Netherlands

Abstract: Eve was the first human being who wanted to know and, what's more, she spontaneously wanted to share her knowledge. So, the idea of Open Access is not new! But it was the internet and more specifically Tim Berners Lee's invention of the World Wide Web that would enable the global materialization of this concept. Unfortunately this does not mean that Open Access will unroll itself automatically. Funders, authors, libraries, publishers, universities and service providers together constitute the research life cycle and all must metamorphose. The original library approach when setting up repositories, 'we-build-it-and-they-will-populate-it', does not seem to work. Open Access goes beyond just introducing new technology. It involves a paradigm change, from thinking in terms of paper snippets that can be owned to delivering identifiable contributions to open knowledge networks.

Keywords: open access, digital publishing, academic libraries, scholarly publishing, publishers, new roles for librarians

Open Access: definition and stakeholders

Open Access is access to research based publications that are peer reviewed, permanently and promptly accessible without payment and re-usable under the sole condition of crediting the author(s). Open Access (OA) involves all the stakeholders of the research life cycle (*figure 1*). To begin with, authors who are the initial copyright holders of the publications. Then the publishers

who organize the peer review. It are libraries who build the national and institutional infrastructures that facilitate permanent and prompt access. Research funders define conditions for granting financing; mandating access to the resulting publications may be among these conditions. Service providers stimulate the reuse of scientific and scholarly findings. Ultimately, legislators may set rules for access to knowledge in our knowledge permeated democracies.

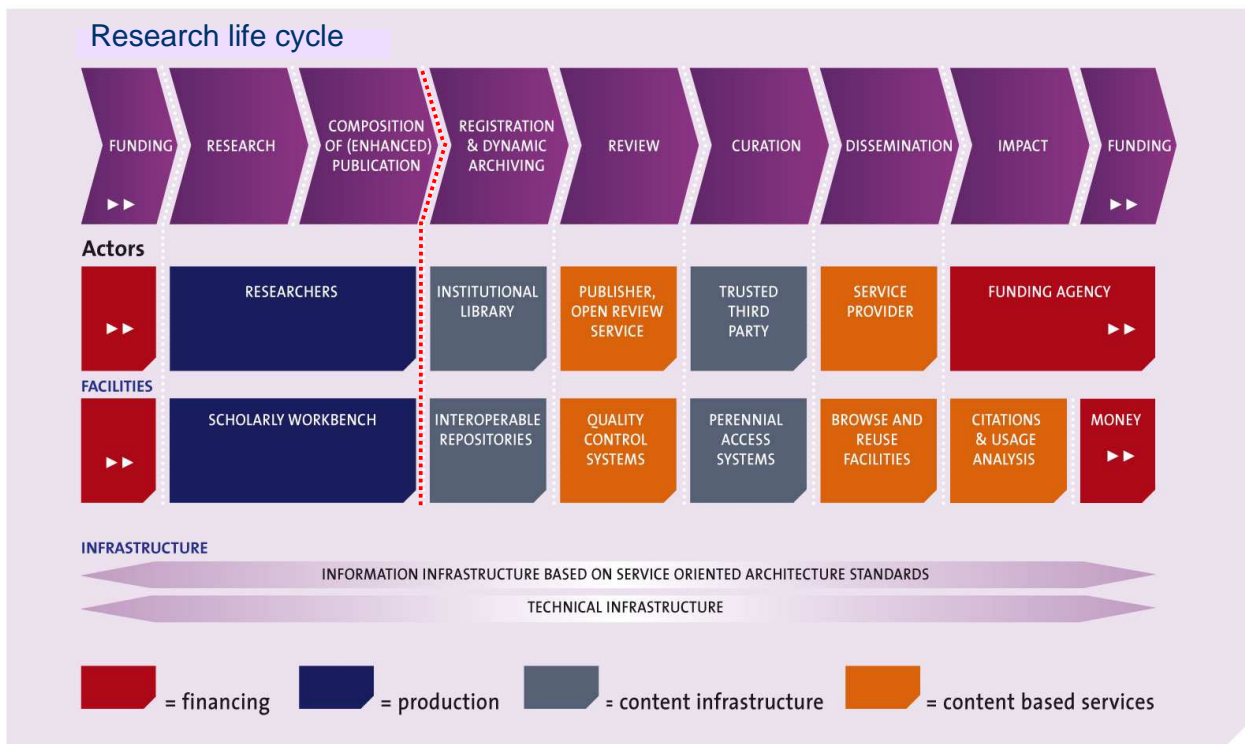


Figure 1 — The research life cycle.

Motives

The arguments in favour of Open Access reflect the stakeholders' diversity. The list of motives includes but is not limited to:

1. Corporate Social Responsibility

More and more HE organisations see OA as part of their 'Corporate Social Responsibility' (= the view that socially responsible behavior is rewarding for an enterprise or organisation). A clear statement of this is found in the Berlin Declaration¹, "*Our mission of disseminating knowledge is only half complete if the information is not made widely and readily available to society.*" which has been signed by more than 250 HE institutions.

2. Citation advantage

Citations expose the talents of both author and institution. The debate on the citation advantage of OA² has been extensive. It is now generally accepted that OA articles generate more citations than non-OA articles in the beginning of their life. Whether this will be the case for their whole life time is still too early to judge as OA articles are still too young for that.

3. Science advancement

"*If I have seen further it is by standing on the shoulders of Giants*" (Newton)

The assumption that a wider circulation of academic publications is beneficial to the progress of science is commonplace. In the forthcoming era of enhanced publications – i.e. publications that include text, data sets, algorithms, visualisations, graphs etc. – various access terms to the components of such publications would make them impractical.

4. Abuse protection

(1) Plagiarism from openly accessible documents can be much easier detected than from texts with restricted access³. (2) In an OA world copyright piracy does not exist; one cannot steal oxygen from the atmosphere.

5. Research transparency

Open access to publications, including data sets, algorithms etc., increases the transparency of research. (1) It enables a broader audience to read the articles critically and try to reproduce the reported results in case of doubt. Ultimately this is the best quality control. (2) It informs a broader audience about the research progress thus reducing the risk of unnecessary and costly duplication.

6. Cheaper distribution

The OA distribution process needs no complicated administration (subscriptions) or access control (licences, passwords, etc.). In the conventional subscription/ licences world the e-only distribution counts for 20% to 25% of the total costs. These costs can be saved globally. It does not mean that these cost savings are spread evenly over the academic institutions.

A brief history of OA

The second world war demonstrated the results of research based warfare (missiles, radar, encryption, atomic bomb). If science could contribute so much to war, why not to peace as well? The idea of a knowledge based society was born. In both 'Western' and 'Eastern' countries higher education exploded in their reciprocal race for supremacy. Research communities became anonymous and publishing grew into an industry including branding and ranking of journals based on the controversial use of citations. The conventional business model was based on exclusive assignment of copyrights to the publisher giving them a monopoly over the control of access to the publications. Thus, they could define the price of access unilaterally which resulted in price increases of 10% to 12% annually over a couple decades. Publishing tycoons like Robert Maxwell (Pergamon Press) and Pierre Vinken (Elsevier) candidly promoted their journals as successful profit machines. As a consequence libraries could no longer afford their subscriptions and had to cancel more and more journals. This in turn gave rise to further price increases as publishers wanted to maintain their revenues. This spiral became known as 'the journals crises'. Eventually the spiral was broken by a new pricing model in the beginning of the century. Publishers halved their price increases in exchange for non-cancellation clauses in multi-year contracts with the libraries, so called licences. This mitigated the problems but left libraries with annual price increases that still exceed their budget growth (if at all) and no freedom to tune their collections to their clients' needs. Publishers could maintain their profits mainly because they could reduce their costs with the help of massive application of information technology. The publishing industry saw a process of business acquisitions resulting in a limited number of dominant players. This process is still ongoing. E.g. in 2004 Springer acquired Kluwer Academic; in 2006 Wiley and Blackwell merged, next to a continuous stream of smaller take-overs.

¹ Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities. <<http://oa.mpg.de/openaccess-berlin/berlindeclaration.html>>.

² OpCit project. The Open Citation Project - Reference Linking and Citation Analysis for Open Archives. <<http://opcit.eprints.org/oacitation-biblio.html>>.

³ See plagiarism detection tools on Shambles: Plagiarism tools <<http://www.shambles.net/pages/staff/ptools/>>.

Since early nineties libraries are questing for a new publishing approach that should no longer be based on copyright monopolies in the hands of publishers. Their hope was based on information technology. Understandably, publishers have paralleled libraries' ambitions with a strong defence of their prevailing business model that has served them so well. At times, the fight has been bitterly. Only recently an appeal to moderate the OA debate was released by the International Federation of Library Associations and the International Publishers Association, IFLA and IPA.⁴

Libraries and OA

The Open Access Initiative Protocol for Metadata Harvesting – OAI PMH 2.0 – of June 2002 meant a breakthrough for libraries in their strive for alternative approaches of giving access to research publications. First technically. OAI divides the world of scholarly and scientific information in two layers: a data layer and a services layer. The data layer is constituted of repositories (= archives), mainly institutionally based, that comply with the OAI standard. They contain the metadata of research publications, generally in the so called DC format. These metadata link to openly accessible research publications that may or may not be contained in the repository. Any service provider applying a OAI harvester can easily collect the metadata and give access to them via a portal. The services may aim at specific target groups e.g. students, subject or disciplinary based communities, professions etc. Web 2.0 facilities like feeds, social tagging or annotating can be offered.

The two-layer approach also holds business-wise. A repository is typically part of the institutional and national infrastructure linked to e.g. the research information system of the institute and the long term preservation facility of the national library. It is designed for machine-machine interaction, needs a concise application of standards, is supply based and belongs to the realm of international cooperation. In short, a repository is a public domain thing. Contrarily, the services in the second layer are end-user oriented activities, demand driven, scalable and bespoke. They operate in competition and are financed on a usage basis. A service is a market thing.

Worldwide academic libraries have embraced the setting up of repositories. The Registry of Open Access Repositories⁵ lists almost 1400 of them but there may be many more. Operationally repositories are not so much different from digital stacks. Both are databases that are part of the institutional infrastructure and containing metadata c.q. catalog cards that refer to documents. In both cases the institutes pay for them and publishers may

restrict access to the documents. Functionally however their role is more or less opposite; stacks give local users access to publications from elsewhere, repositories give external users access to locally produced publications. The migration from the classical to the new function requires a mental paradigm shift. The classical stacks holding library is a local monopolist defining their own collection, cataloging rules, etc. The repository holding library is a small building block in a global knowledge infrastructure based on common standards and protocols. The 'they-have-to-cooperate-with-us' attitude must change in a 'we-have-to-cooperate-with-them' approach. Intra-operability has to change in inter-operability. This is not always easy for libraries.

Libraries are also active in the second OAI layer, the services. They often build the research window of their university and many feed the individual web sites of their professors with repository content. In more and more institutes the repository in combination with the institutional research information system forms the basis for the institutional annual report and research assessments.

In how far all these activities make access to knowledge cheaper, for libraries the initial OA driver, remains to be seen. Expectations are that Open Access will turn out to be cheaper for the academic community as a whole because the significant costs for contracts and shielding in the subscription/licences model will be eliminated. Moreover, the Open Access model is based on the operation of market forces, while the traditional model is based on copyright monopolies.

At institutional level, it is relatively simple to carry out a cost comparison between the OA model and the conventional model. In the Open Access model, the institution pays a publication fee – which differs from publisher to publisher – for all its articles. Price lists for publication fees can be found on the web sites of Sherpa/RoMEO⁶ and BioMedCentral.⁷ Moreover, more than 50% of the OA publishers do not charge a fee at all.⁸ The calculation can take account of the fact that the bodies that finance research are increasingly viewing publication as the culmination of the study concerned and therefore accept the publication costs as forming part of the research budget. The Sherpa/JULIET⁹ Web site gives an overview of the policies of the major research financiers.

⁴ IPA. *International Publishers and Librarians Agree to Enhance: The Debate on Open Access* <http://www.nuv.nl/SiteCollectionDocuments/PR_on_open_access_statement_20090520.pdf>.

⁵ ROAR <<http://roar.eprints.org/>>.

⁶ Sherpa/RoMEO. *Publishers with Paid Options for Open Access*. <<http://www.sherpa.ac.uk/romeo/PaidOA.html>>.

⁷ BioMedCentral. *Comparison of BioMed Central's Article Processing Charges with those of other publishers*. <<http://www.biomedcentral.com/info/authors/apccomparison/>>.

⁸ ALPSP. *The Facts About Open Access*. <http://www.alpsp.org/ngen_public/article.asp?id=200&did=47&aid=270&st=&oaid=-1>.

⁹ Sherpa/Juliet. *Research funders' open access policies*. <<http://www.sherpa.ac.uk/juliet/index.php>>.

Established in 1952, the **Institute of Social Studies** in the Hague is an international graduate school with typically 400 students per year. Its research programme results in books, reports and about 60 articles in peer reviewed journals annually. An intangible open access advantage is the free access that ISS alumni, who are often based in developing countries, will have to these articles.

Open Access		Subscriptions	
Publication fees ISS	€ 120,000	Subscriptions & licences	€ 180,000
		Document supply + copyright clearance fees	€ 65,000
Platform or aggregation costs	€ 25,000	Collection management	€ 60,000
	€ 145,000		€ 305,000

- 60 peer reviewed articles @ 2000. The publication fee is an estimate based on the pricelist published by BioMedCentral. The list refers to STM journals mainly (<http://www.biomedcentral.com/info/authors/apccomparison/>). For social sciences
- ISS assumes that even in a full open access world still some aggregation or platform fees will be needed
- Includes out of pocket costs to third parties (other libraries, publishers) and in house handling costs.
- This is a rough estimate. Includes defining the collection, acquisitions and administration, shielding access and copyright

September 2008.
Michel Wesseling,
Head of the Office of Library and Information Technology Services.

Table 1 — Cost comparison Institute of Social Studies.

Established in 1636, today **Utrecht University** is an internationally renowned top research university with almost 30.000 students and over 8.500 staff. Annually UU published 5000 articles in peer reviewed journals; of these 1500 result from externally funded research (figures 2005). Utrecht University is a signatory of the Berlin Declaration on Open Access.

Open Access		Subscriptions	
Publication fees UU ¹	€ 6,500,000	Subscriptions & licences	€ 4,500,000
		Document supply + copyright clearance fees + collection management	€ 1,000,000
	€ 6,500,000		€ 5,000,000
Publication fees research funders ²	€ 1,950,000		
Remaining costs UU	€ 4,550,000		

- 5000 peer reviewed articles @ 1300
- 1500 peer reviewed articles @ 1300
- Includes out of pocket costs to third parties (other libraries, publishers) and in house handling costs.
- Includes defining the collection, acquisitions and administration, shielding access and copyright issues.

September 2008.
Bas Savenije,
University Librarian of Utrecht University.

Table 2 — Cost comparison Utrecht University.

In the traditional model, an institution not only pays subscription and/or licensing fees but also reproduction fees, charges for individual articles that are requested (via interlibrary loan or directly from the publisher), and the contract costs already referred to. The latter are difficult to calculate, but they are nevertheless considerable. The library must first of all determine its purchasing policy; this often involves time-consuming consultations with faculty. The accessions then need to be registered in a complex system of subscription records or laid down in licences; these are legal documents, specifically regulating access rights, that are by no means trivial and usually only drawn up after lengthy negotiations. The institution must then implement the agreed access restrictions by means of shielding constructions such as IP addresses, passwords and proxy servers, or special software such as SFX. Finally, these restrictions mean that filling the institution's repository is a laborious matter and therefore needlessly expensive.

The calculations will show that the cost advantages of the Open Access model are not distributed equally between institutions. In the age of Open Access, genuine research universities may even find themselves having to pay more than in the age of the traditional model. Such institutions also need to spend more (than for example institutions focusing on teaching) on other research facilities such as laboratories, supercomputers and grids, and on attracting top researchers. Indeed, noblesse oblige. Estimates of the costs involved in publishing research results using the Open Access method range from 1.5% to 2% of the research budget.

Recently two Higher Education institutes in the Netherlands have done such a calculation (see tables 1 and 2). As expected, open access is very beneficial to the Institute of Social Studies with its focus on an educational programme. The outcome for Utrecht University, a top research institution, is surprisingly more nuanced. If research funders are prepared to pay for the publication fees for articles that result from their projects, open access is also financially advantageous to this research university.

One pressing problem has become the phase transfer from A to B. An institution that decides to use Open Access does not immediately cease having to pay its subscription and/or licensing fees. In fact, it pays not only for publishing its own articles but also – through the subscription or licensing fees – for publication by institutions that have not yet made the switch. This effect can be prevented if large groups switch together, although that requires co-ordination. One example is CERN's SCOAP3¹⁰ Project. This involves CERN – acting on behalf of the high-energy physics community – defining the conditions for quality control and open distribution of articles and requesting publishers to submit quotations. This discipline is consequently switching to Open Access all in one go. Other examples are the Dutch university library consortium UKB and the Max Planck Gesellschaft

¹⁰ SCOAP3 <<http://scoap3.org/>>.

which in experiments with Springer¹¹ could make the switch to Open Access publishing for all their authors in Springer journals. Recently the University of California concluded a similar deal.¹² The economist Ted Bergstrom has already referred, as far back as 2001, in a now famous parable,¹³ to the need for such co-ordination. However, this is not the strongest point of libraries with their long tradition of local dominance.

The essential point of the Springer deal is that universities do not have to pay extra during the transition period for having the articles of their researchers published in Open Access in a Springer journal. *Box 1* gives a good insight in the effects of the contract between the Dutch library consortium UKB and Springer for a university i.c. Wageningen University and Research Centre.

Springer Open Choice also Free in 2008 and 2009

For the time being, Wageningen UR authors do not yet have to pay for Springer Open Choice¹⁴. In 2007, the library cooperative UKB and publisher Springer signed a declaration of intention on Open Access. The agreement stated that all publications from authors associated with a Dutch university and published in a Springer journal in 2007 would be made accessible through Open Access free of charge. In 2007, Dutch universities published 1162 articles in Springer journals. In the meantime, this pilot has been expanded to a maximum of 1250 articles for 2008 and 2009. After reaching this quota, \$1500 per article will have to be paid. If an external subsidizer pays for the admission to Open Choice, this publication will not count towards the maximum number of articles. Articles are entered in Open Choice in the order in which they are received.

If an author has submitted an article to a Springer journal and the article is accepted, he/she receives a screen with the notice of acceptance that explains the complimentary admission to Open Choice. The screen also allows the author to register the publication for Open Choice. The university libraries will also be notified if an article has been accepted, so that they can register it in the Institutional Repository (WaY¹⁵).

Box 1 — The effects of the UKB-Springer deal for Wageningen UR.

¹¹ *Together Again: Springer, Max Planck Agree To New "Experimental" Deal.* <<http://www.libraryjournal.com/info/CA6528977.html>>.

¹² *ANDERSON I. UC libraries and Springer sign pilot agreement for open access journal publishing.* <<http://www.universityofcalifornia.edu/news/article/19335>>.

¹³ *BERGSTROM T. The Parable of the Anarchists Annual Meeting* <<http://www.econ.ucsb.edu/~tedb/Journals/anarchists.pdf>>.

¹⁴ *Springer. Springer Open Choice.* <<http://www.springer.com/open+access/open+choice?SGWID=0-40359-0-0-0>>.

¹⁵ *Wageningen Yield* <<http://library.wur.nl/way/>>.

Publishers and OA

In order to understand how Open Access publishing works compared to its classical variant, we might do worse than take a brief look at the publication process of scientific articles (*figure 2*). The quality control in both approaches is identical: a pre-print goes through an editorial and peer review process that, often after revision of the manuscript, may result in a post-print. Then, before proceeding to publication of the article, publishers expect to be remunerated for their work, namely the organisation of the quality control and subsequent dissemination of the article via one of their, sometimes highly branded, journals. At this juncture, the two approaches bifurcate. In the classical approach publishers require exclusive, complete and irreversible assignment of copyright which gives them control over access to the article. Subsequently, they transubstantiate this control into revenue via subscriptions, licences, copyright clearance fees and various conditions of reuse. Conversely, Open Access publishers require a sum of money upfront, the so-called 'publication fee' or 'author fee'. Upon its receipt, they publish the article promptly in an open access Internet journal where the sole condition of reuse is the correct attribution of the content to its author(s).

The copyright monopoly that conventional publishers acquire is used for various purposes. The main use is setting the access prices of their journals without the fear of competition. It has led to price lists that show no relation to quality (measured in terms of impact) or quantity (measured) in terms of volume. Publishers may also use copyright ownership to define the reuse conditions of the articles e.g. inclusion in course packs, posting in repositories, usage for interlibrary document supply and

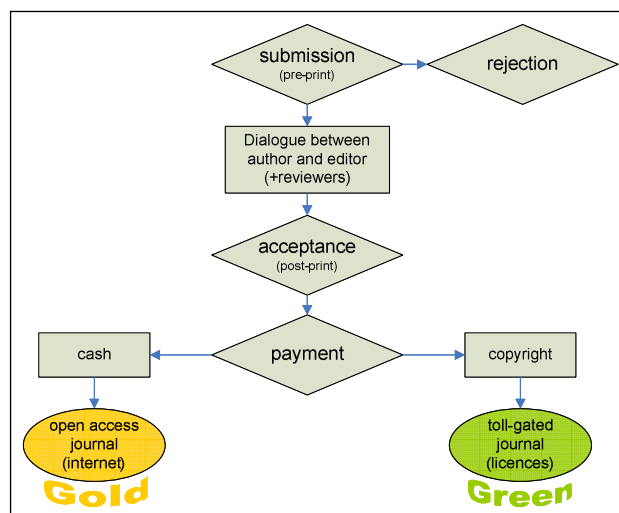


Figure 2 — Open Access and Toll Gate publishing model.

translations. They may withdraw articles¹⁶ from their web site or re-publish them in fake journals.¹⁷

The conventional publishers have defended their business model at many occasions culminating in the PRISM campaign¹⁸ of the American Association of Publishers in 2007 claiming that OA would destroy the peer review system and that public access equals government censorship. Although there has been a recent appeal of IFLA and IPA to moderate the OA debate the anti OA publishers lobby is still on. Current arguments are that OA will create job losses in the conventional publishing industry and that during a transition period the co-existence of two models (Toll Gate and Open Access) will be more expensive to institutes.

Nevertheless the number of Open Access publishers and Open Access journals is growing. The Directory of Open Access Journals DOAJ¹⁹ shows over 4000 journals and as of Oktober 2008 the Open Access publishers do have their own association OASPA²⁰.

Funders and OA

Several research funders and also some universities mandate open access²¹ to the publications of the research they fund and are prepared to pay for the publication fee. A typical example is the European Commission with their Open Access Pilot in the Seventh Research Framework Programme F7.²² FP7 is the Commission's €50 billion research programme, running from 2007 to 2013. The Pilot covers approximately 20% of the budget and applies to the following areas: Energy, Environment, Health, ICT, e-Infrastructures, Science in Society, Socio-economic Sciences and Humanities. The Commission requires Open Access to all peer-reviewed articles resulting from research in these areas within six or twelve months of publication. If we estimate that every €100,000 of research funding results in one peer-reviewed article, the FP7 pilot will provide open access to something like 100,000 articles. All articles must be deposited into institutional or subject-based repositories.

In order to enable authors to meet these publishing requirements, the Commission reimburses the publication fee for articles published in Open Access journals. The Commission has also tendered for the provision of a pan-European repository²³ where authors who do not have an institutional or subject-based repository at their disposal can store their article.

Currently, Open Access journals represent some 15% of the market of scientific and scholarly publishing. Statistically, this means that 15,000 articles of the FP7 pilot meet the requirements of the Commission. As copyright owners, their authors can deposit these articles in a repository immediately upon their publication.

The remaining 85,000 articles have to follow the Green road to Open Access. They still have to be published in a subscription-based journal, which involves assignment of copyright to the publisher. And then the authors "*should make their best efforts to negotiate copyright and licensing conditions that comply with the Open Access pilot in FP7*" says the Commission in its Open Access Pilot leaflet²⁴. Authors can discover exactly what this means by going to the RoMeo/SHERPA²⁵ website. To help authors the Commission has published model amendments to publishing agreements²⁶ plus accompanying cover letters to their publisher in all European languages.

Yet, the Commission's main supportive step is still missing. One of the standard mechanisms for the EC to realise its policies is putting out tenders and calls for bids or proposals. Indeed, the Commission has tendered the provision of a pan-European repository for orphan authors in FP7. But lacking an appropriate peer review system, the Commission sends authors off to find their way around the RoMEO/Sherpa labyrinth. Instead, it could have tendered for bids to establish adequate non-proprietary review systems. Such a measure would follow up the policy advice of the important Study on the Economic and Technical Evolution of the Scientific Publication Markets in Europe²⁷ which it published three years ago. Such an invitation to tender (ITT) might run as follows:

¹⁶ FOSTER A.L.. *Elsevier's Vanishing Act*
<<http://chronicle.com/free/v49/i18/18a02701.htm>>.

¹⁷ OCKERBLOOM J.M. *What you're asked to give away.*
<<http://everybodyslibraries.com/2009/05/08/what-youre-asked-to-give-away/>>.

¹⁸ PRISM <<http://www.eurekajournalwatch.org/index.php/PRISM>>.

¹⁹ DOAJ <<http://www.doaj.org/>>.

²⁰ OASPA <<http://www.oaspa.org/>>.

²¹ Sherpa/JULIET <<http://www.sherpa.ac.uk/juliet/>>.

²² European Commission. *Research. Open access pilot in FP7.*
<<http://ec.europa.eu/research/science-society/index.cfm?fuseaction=public.topic&id=1680>>.

²³ MORAIS PIRES C. *INFRA-2009-1.2.3*
<ftp://ftp.cordis.europa.eu/pub/ftp7/ict/docs/e-infrastructure/20081216-infoday-morais_en.pdf>.

²⁴ European Commission. *Open Access Pilot in FP7.*
<http://ec.europa.eu/research/science-society/document_library/pdf_06/open-access-pilot_en.pdf>.

²⁵ Sherpa/RoMEO. *Publisher copyright policies & self-archiving. Statistics for the 626 publishers on this list* <<http://www.sherpa.ac.uk/romeo.php?stats=yes>>.

²⁶ European Commission. *Research. Open access pilot in FP7.*
<<http://ec.europa.eu/research/science-society/index.cfm?fuseaction=public.topic&id=1680>>.

²⁷ European Commission. *Study on the Economic and Technical Evolution of the Scientific Publication Markets in Europe*
<http://ec.europa.eu/research/science-society/pdf/scientific-publication-study_en.pdf>.

"The European FP7 Programme (€50 billion, 2007-2013) will generate a stream of published research. For seven areas of research the Commission requires that such published research appears as openly accessible peer-reviewed articles. In order to have appropriate review procedures in place to process these articles, the Commission plans to tender the reviewing process under the following conditions:

1. The reviewing process must be independent, rigorous and swift.
2. The reviewing may be anonymous, named or open (to be negotiated).
3. As a result of the reviewing process, articles will be marked 1 to 5.
4. For articles marked 3 to 5 adequate Open Access publication platforms must be offered (e.g. new Open Access journals). Alternatively, authors may publish their articles in any existing OA journal.
5. In review procedures the Commission will weigh articles with marks 3, 4 and 5 as if they were published in journals with impact factors 1-3, 4-8 and 9-15 respectively (These figures are nominal and subject to disciplinary calibration).
6. Upon publication all articles will be deposited in a certified (institutional) repository.

A yield of at least 100.000 articles across the selected research areas is anticipated during the Programme period.

It is possible that more than one proposal will be accepted in order to cater for different disciplines. The Commission will seek the advice of EURAB/ESF/EUA/EUROHORCS/ERC in the selection process.

Box 2 — A possible Invitation to Tender for Non-proprietary peer review systems

The wording of the proposed IIT above requires some explanation:

- This imaginary IIT is geared to the EC but could easily be adapted to (consortia of) funders or universities alike which have developed distinct Open Access policies.
- Condition 4 above creates an adequate Open Access solution for all authors within the mandate. Of course, if authors insist on publishing their article in a Toll Gate journal (and undertake the complicated process of negotiating delayed open access to their post-print) they may still do so.
- The announcement that in review procedures articles with marks 3 to 5 will be weighed as if they were published in journals with certain impact factors aligns this approach with the existing review systems. The advice of prestigious European research authorities – all fervent Open Access advocates – authorises this important step. This nuanced approach also disincentivises a costly iteration of rejection and re-submission of articles.
- Of the total of 100,000 articles, 15,000 can already be published in Open Access journals today. So there is a market for 85,000 open access articles e.g. for about 120 new Open Access journals (120 journals × 7 years × 100 articles/journal-year = 84,000 articles.)
- Proposals may include existing Open Access journals. Especially for young Open Access journals, i.e. journals

that have not yet gathered a journal impact factor, this call offers the opportunity to gain accreditation by the Commission.

- Classical publishers could seize the opportunity to switch Toll Gate journals to Open Access. The expected number of articles should stimulate the Open Access market.

Authors and OA

Authors of rich institutes have always lived in an open access environment 'avant la lettre'; to them both publishing and reading were free. So, originally they did not have a natural inclination to OA. Only very few early birds like Andrew Odlyzko and Stevan Harnad recognized the flaws of the classical publishing system in a digital world and advocated new approaches. Gradually others followed. In 2001 38.000 authors in the biomedical disciplines signed the PLoS open letter.²⁸ In 2007 27.000 people from academia signed the EC petition²⁹ (more than 700 on behalf of their organisation). Studies by Alma Swan³⁰ demonstrated author preparedness to comply with Open Access mandating.

The main motive for authors to be in favour of OA is their wish to see their publications be circulated and reused as widely as possible. In this wish personal and ethical reasons concatenate. Open Access generates more citations and thus contributes to status and prestige of the author which in turn benefits their career and the granting of their research proposals. But Open Access also recognizes the argument that most research is financed by public money. So it would be unethical to restrict access to its results to the employees of rich institutions only.

Authors may experience obstacles for publishing their articles in Open Access:

- (1) There is no attractive OA journal available. Attractiveness can both relate to the subject and the impact factor of the available journals. These concerns are realistic although both the disciplinary coverage and the impact factors of OA journals have grown considerably over the last couple of years. The Directory of Open Access Journals demonstrates a stable growth of more than two new OA journals per day and the impact factors of a number of OA journals are excellent.
- (2) If there is no appropriate OA journal authors can publish their article in a conventional journal that requires the assignment of their copyrights. Then it is the publisher who defines the access conditions of the article. About

²⁸ PLoS. *Public Library of Science. Read the Open Letter.* <<http://www.plos.org/support/openletter.shtml>>.

²⁹ *Petition for guaranteed public access to publicly-funded research results* <<http://www.ec-petition.eu/>>.

³⁰ SWAN A. and BROWN S. *Open access self-archiving* <<http://eprints.ecs.soton.ac.uk/10999/>>.

63%³¹ of the conventional journals allow archiving of the post-print version of the published article in the institutional repository and permit open access to this manuscript after an embargo period. This route to open access goes under the name Green route. It is quite complicated as it needs publisher's permission, involves extra administrative work and citations of post scripts are not included in citation indexes like Web of Science and Scopus. In fact the Green route is a form of parallel publishing that requires the maintenance of the underlying subscription/licences publishing model. Ultimately it is unsustainable but for the time being it may serve a cause. Proponents of the Green route claim that after it has reached a certain critical mass publishers will give in, leave the subscription model and step over to the Open Access model.

In the meantime there is a growing number of authors that play the "Don't ask, don't tell" game³². They still publish in conventional subscription journals and at the same distribute their articles online as if they were published in Open Access thus sometimes violating their copyright contracts.

Practically all issues about copyrights in academia are speculative. In general the opening phrase of a discussion reads, "Would it be allowed to...?". The only one who can decide is the owner of the copyrights and there are two methodologies to find out their answer: (1) asking, (2) testing.

Ad 1. Quite a few people in academia are inclined not to ask the copyright owner about copyrights but someone else e.g. a lawyer or a librarian. This leads to abstract, elusive or fuzzy third party guesses that in turn provoke others to react upon thus creating the perfect soil for committees, seminars, reports etc. So, my advise is to approach the copyright owner – i.e. the author or the publisher - and ask them. Or negotiate with them if you are in a position to do so. That's all.

Ad 2. When a situation is less clear testing it is not uncommon in an academic environment. Why not applying this methodology also in opaque matters of access to or reuse of publications? Do what you think is well, be candid about it and observe what happens. Sometimes people seem to be frightened to do so. Indeed, scared people limit their own scope and bold people extend it. But this is a fact of life and has nothing to do with copyrights. Here also, committees, seminars and reports are no remedy.

However, it could be interesting to create a database of relevant facts where authors can register concrete copyright practices like "My publisher (name) accepted the following copyright statement" or "I did this and then that happened" thus sharing their experiences.

³¹ EPrints. *Journal Policies - Summary Statistics So Far*. EPrints <<http://romeo.eprints.org/stats.php>>.

³² [Shieber S.M.] "Don't ask, don't tell" rights retention for scholarly articles. <<http://blogs.law.harvard.edu/pamphlet/2009/06/18/dont-ask-dont-tell-rights-retention-for-scholarly-articles/>>.

Above all, authors should realise that the only legal right that a publisher really needs is the right to publish their manuscript. All other rights that the author assigns to the publisher are gifts in kind that boost the economic position of the publisher, especially when the transfer is an exclusive one. There are a number of copyright statements that limit the transfer of rights to a publisher to what is necessary for publication, for example JISC-SURF's Licence to publish³³, SPARC's Author Addendum³⁴ or the 'Model amendment to publishing agreement'³⁵ of the European Commission (available in all European languages). Some prestigious scientific journals have begun by themselves to provide authors with standard licences to publish that do not require authors to assign their copyright to the publisher e.g. Nature³⁶ and Science³⁷.

In the meantime the character of publications is metamorphosing as authors more and more seem to understand and apply the full potential of digital publishing and make data sets, algorithms, videos, blog entries and even post-publication reviews part of their (dynamic) article. New standards for the structuring and exchange of these compound or enhanced publications are developed by Herbert van de Sompel and others under the name OAI-ORE,³⁸ Open Access Initiative – Object Reuse and Exchange. Commercial parties like Microsoft are building Technologies for the Scholarly Communications Lifecycle³⁹ virtual research environments with tools for co-operating, authoring, publishing and archiving. The underlying paradigm is that the components of such enhanced publications are autonomous open web resources glued together by rdf-triples and presented in a resource map. Unlike hyperlinks the rdf-triples are two-way links that describe the relation between the linked objects in a short object-predicate-subject sentence. Thus the linked object 'knows' that it is linked to by 'whom' and 'why'. Citation indexes become an inherent aspect of web resources and a step forward is made toward the semantic web.

Needless to say that this approach becomes unworkable when all these web resources will have their own access

³³ JISC-SURF. *Copyright toolbox* <<http://copyrighttoolbox.surf.nl/copyrighttoolbox/>>.

³⁴ SPARC. *Author Rights: Using the SPARC Author Addendum to secure your rights as the author of a journal article* <<http://www.arl.org/sparc/author/addendum.shtml>>.

³⁵ European Commission. *Research. Open access pilot in FP7*. <<http://ec.europa.eu/research/science-society/index.cfm?fuseaction=public.topic&id=1680>>.

³⁶ Nature. *License to publish* <http://www.nature.com/authors/editorial_policies/license.html>.

³⁷ Science. *License to Publish* <<http://www.sciencemag.org/feature/contribinfo/prep/license.pdf>>.

³⁸ OAI-ORE. *Open Access Initiative - Object Reuse and Exchange*. <<http://www.openarchives.org/ore/>>.

³⁹ Microsoft Research. *Technologies for the Scholarly Communications Lifecycle*. <<http://research.microsoft.com/en-us/collaboration/focus/education/scholarlycomm.aspx>>.

limitations. Where Open Access may be desirable for authors when it comes to more citations, it becomes imperative in the new publishing paradigm – ‘Publishing 2.0’ – that comes with e-science (*figure 3*, taken from Understanding the Semantic Web⁴⁰)

Service providers and OA

OA services range widely, from mini – e.g. personal web sites – via meso – e.g. institutional web sites, publisher web sites and disciplinary community portals – to macro services like Google Scholar⁴¹, BASE⁴², OAIster⁴³ or Europeana⁴⁴. An special mini-service is Publish or Perish⁴⁵ supplied by Australian Anne Will Harzig. It collects all the articles of an author in Google Scholar and automatically generates a number of indexes like the h-index for that author. Google Scholar itself would become more useful to Open Access if it would create a subset (“Google Scholar Open”) that only generates hits that give open access to the full text. An interesting meso-service is supplied by eIFL⁴⁶, an organisation that stimulates transitional and developing countries to participate in the OA movement.

Next to OA based end user services, services for OA professionals are emerging. PKP⁴⁷ delivers widely used open source journal systems, conference systems and harvesting tools. The Open Access Scholarly Information Sourcebook⁴⁸ supplies an overview of the current developments on practically all relevant OA issues.

DRIVER⁴⁹ is a European project that services both end users and professionals. For end users it enables searching of the content of over 200 repositories, with approximately 1M documents, not only in the usual way (per author, keyword, document type, publication date or personal profile) but also per repository, country, language, community, collection or subject.

For professionals DRIVER offers: General support (incl. a wiki⁵⁰ and a FAQ), a Harvesting service⁵¹, Open source software (under the name D-NET⁵²), Mentor service⁵³ for novices in the repository world (including a tutorial⁵⁴), Metadata guidelines⁵⁵ reflecting the agreements that must be made to enhance the interoperability of the various repositories, a Validator⁵⁶ that test and reports about the repository compliance with the OAI-PMH protocol and the metadata guidelines and finally Studies⁵⁷ on business models, OAI standards, copyrights and an inventory of repositories in the EU⁵⁸.

To conclude

In order to be successful OA needs a robust interoperable **infrastructure** and user focussed **services**. For most OA services a certain level of **coverage** is also a critical condition for success.

Although OA needs an integrated approach of practically all stakeholders in the research life cycle their priorities may be different. The infrastructure typically lies in the domain of the co-operating (national and institutional) libraries. Coverage can be stimulated by funders mandating parallel publishing (Green OA). Not only the national Research Council(s), but also universities, charities, applied sciences and technology funders and even companies are relevant funders in this respect. Ultimately, non-proprietary peer review services (as offered by e.g. OA publishers) are key to the academic community and to society. The publishing market could and should be stimulated to offer these services e.g. by tendering them.

End user OA services are easily offered with the help of modern ICT. The main question is are they demanded. Novice OA service providers, e.g. libraries, should include the sustainability of a service before they start one.

⁴⁰ SPIVACK, Nova. *Understanding the Semantic Web and Twine Talk*. <<http://www.slideshare.net/novaspivack/nova-spivack-understanding-the-semantic-web-and-twine-talk>>.

⁴¹ Google Scholar <<http://scholar.google.com/>>.

⁴² BASE. *Bielefeld Academic Search Engine* <<http://base.ub.uni-bielefeld.de/en/index.php>>.

⁴³ OAIster <<http://www.oaister.org/>>.

⁴⁴ Europeana <<http://www.europeana.eu/portal/>>.

⁴⁵ HARZIG A.W. *Publish or Perish*. <<http://www.harzing.com/pop.htm>>.

⁴⁶ eIFL –OA <<http://www.eifl.net/cps/sections/services/eifl-oa>>.

⁴⁷ PKP <<http://pkp.sfu.ca/>>.

⁴⁸ OASIS <<http://www.openoasis.org/>>.

⁴⁹ DRIVER <<http://www.driver-community.eu/>>.

⁵⁰ DRIVER Wiki <<http://www.driver-support.eu/pmwiki/>>.

⁵¹ DRIVER. *Repository registration*. <<http://validator.driver.research-infrastructures.eu/validatorWeb/pages/registrationInfo.jsp>>.

⁵² DRIVER. *D-NET* <http://www.driver-repository.eu/D-NET_release>.

⁵³ DRIVER. *Mentor Service* <<http://www.driver-support.eu/mentor.html>>.

⁵⁴ DRIVER *Mentor Service [tutorial]* <http://www.nottingham.ac.uk/toolkits/play_470>.

⁵⁵ DRIVER. *Exposing textual resources with OAI-PMH*. <http://www.driver-support.eu/documents/DRIVER_guidelines_10.pdf>.

⁵⁶ DRIVER *Validation* <<http://validator.driver.research-infrastructures.eu/validatorWeb/pages/validatorHome.jsp>>.

⁵⁷ DRIVER *Studies*. <<http://validator.driver.research-infrastructures.eu/validatorWeb/studies.do?action=load>>.

⁵⁸ VERNOOY-GERRITSEN M., PRONK G. and van der GRAAF M. *Three Perspectives on the Evolving Infrastructure of Institutional Research Repositories in Europe* <<http://www.ariadne.ac.uk/issue59/vernooy-gerritsen-et-al/>>.

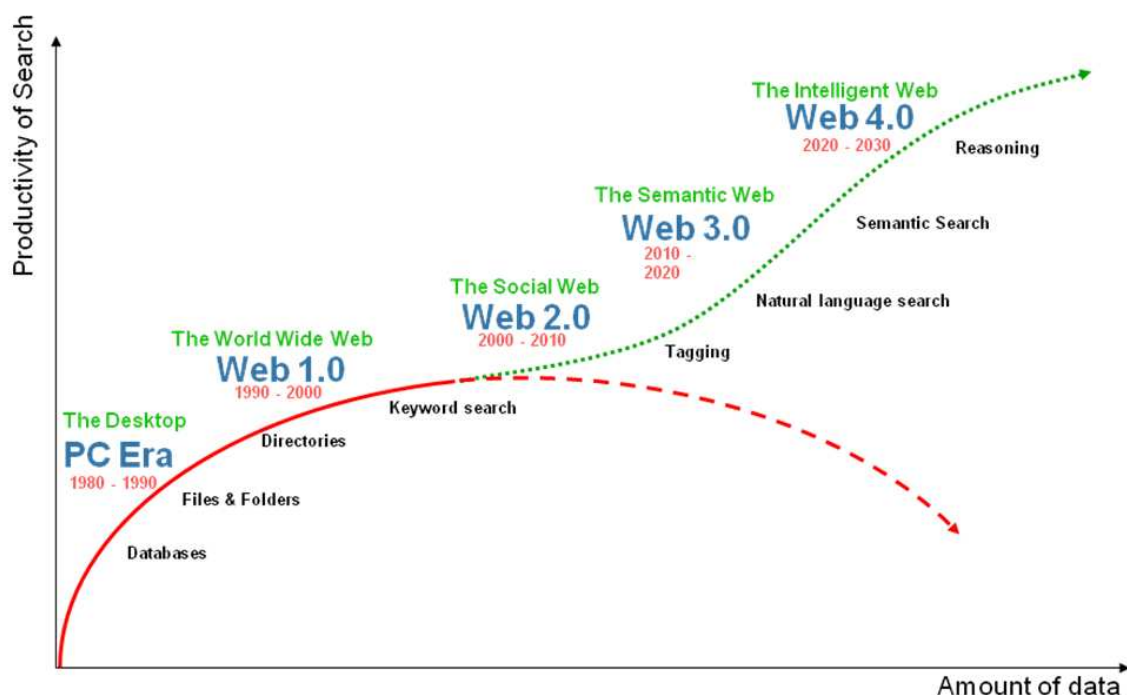


Figure 3 — Developments in searching the web

References

- ALPSP. (The Association of Learned and Professional Society Publishers). *The Facts About Open Access*. [online]. [Cited 2009-06-07]. Available from: <http://www.alpsp.org/ngen_public/article.asp?id=200&did=47&aid=270&st=&oaid=-1>.
- ANDERSON, Ivy. UC libraries and Springer sign pilot agreement for open access journal publishing. *UC Newsroom* [online]. 2009-01-21 [cited 2009-06-07]. Available from: <<http://www.universityofcalifornia.edu/news/article/19335>>.
- BERGSTROM, Ted. *The Parable of the Anarchists Annual Meeting : How can there be monopoly profits in academic journals?* [online]. [Cited 2009-06-07]. Available from: <<http://www.econ.ucsb.edu/~tedb/Journals/anarchists.pdf>>. *Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities* [online]. Conference Open Access to Knowledge in the Sciences and Humanities. 20-22 Oct 2003, Berlin [cited 2009-06-07]. Available from: <<http://oa.mpg.de/openaccess-berlin/berlindeclaration.html>>.
- BioMedCentral. *Comparison of BioMed Central's Article Processing Charges with those of other publishers* [online]. [Cited 2009-06-07]. Available from: <<http://www.biomedcentral.com/info/authors/apccomparison/>>.
- European Commission. *Open Access Pilot in FP7* [online]. [Cited 2009-06-07]. Available from: <http://ec.europa.eu/research/science-society/document_library/pdf_06/open-access-pilot_en.pdf>. European Commission. *Research. Open access pilot in FP7* [online]. [Cited 2009-06-07]. Available from: <<http://ec.europa.eu/research/science-society/index.cfm?fuseaction=public.topic&id=1680>>.
- European Commission. *Study on the Economic and Technical Evolution of the Scientific Publication Markets in Europe* [online] : Final Report – January 2006 [cited 2009-06-07]. Available from: <http://ec.europa.eu/research/science-society/pdf/scientific-publication-study_en.pdf>.
- FOSTER, Andrea L. Elsevier's Vanishing Act : To the Dismay of Scholars, the Publishing Giant Quietly Purges Articles from Its Database. *Chronicle of Higher Education* [online]. 10 Jan 2003 [Cited 2009-06-07]. Available from: <<http://chronicle.com/free/v49/i18/18a02701.htm>>.
- HARZIG, Anne Will. Publish or Perish. [SW online]. [Cited 2009-06-07]. Available from: <<http://www.harzing.com/pop.htm>>.
- IPA. *International Publishers and Librarians Agree to Enhance: The Debate on Open Access*. [online]. Press Release. Geneva/The Hague 20 May 2009 [cited 2009-06-07]. Available from: <http://www.nuv.nl/SiteCollectionDocuments/PR_on_open_access_statement_20090520.pdf>.
- Microsoft Research. *Technologies for the Scholarly Communications Lifecycle* [online]. [Cited 2009-06-07]. Available from: <<http://research.microsoft.com/en-us/collaboration/focus/education/scholarlycomm.aspx>>.
- MORAIS PIRES, Carlos. *INFRA-2009-1.2.3: Scientific Information Repository supporting FP7* [online]. [Cited 2009-06-07]. Available from: <ftp://ftp.cordis.europa.eu/pub/fp7/ict/docs/e-infrastructure/20081216-infoday-morais_en.pdf>.
- Nature. *License to publish* [online]. [Cited 2009-06-07]. Available from: <http://www.nature.com/authors/editorial_policies/license.html>.
- OAI-ORE. Open Access Initiative - Object Reuse and Exchange [online]. [Cited 2009-06-07]. Available from: <<http://www.openarchives.org/ore/>>.
- OCKERBLOOM, John Mark.: What you're asked to give away. *Everybody's Libraries* [online]. 8 May 2009 [cited 2009-06-07]. Available from: <<http://everybodyslibraries.com/2009/05/08/what-youre-asked-to-give-away/>>.

- OpCit project. *The Open Citation Project - Reference Linking and Citation Analysis for Open Archives* [online]. [Cited 2009-06-07]. Available from: <<http://opcit.eprints.org/oacitation-biblio.html>>.
- Petition for guaranteed public access to publicly-funded research results [online]. [Cited 2009-06-07]. Available from: <<http://www.ec-petition.eu/>>.
- PLoS. (Public Library of Science). *Read the Open Letter* [online]. [Cited 2009-06-07]. Available from: <<http://www.plos.org/support/openletter.shtml>>.
- PRISM [Partnership for Research Integrity in Science and Medicine]. Eureka. Science Journal Watch [online]. Last mod. 2007-09-20 [cited 2009-06-07]. Available from: <<http://www.eurekajournalwatch.org/index.php/PRISM>>.
- Science. *License to Publish* [online]. [Cited 2009-06-07]. Available from: <<http://www.sciencemag.org/feature/contribinfo/prep/license.pdf>>.
- SCOAP3 (Sponsoring Consortium for Open Access Publishing in Particle Physics) Towards Open Access publishing in High Energy Physics [online]. [Cited 2009-06-07]. Available from: <<http://scoap3.org/>>.
- Sherpa/RoMEO. *Publishers with Paid Options for Open Access* [online]. Last Updated: 2009-06-11 [cited 2009-06-11]. Available from: <<http://www.sherpa.ac.uk/romeo/PaidOA.html>>.
- Sherpa/RoMEO. *Publisher copyright policies & self-archiving. Statistics for the 626 publishers on this list* [online]. [cited 2009-06-11]. Available from: <<http://www.sherpa.ac.uk/romeo.php?stats=yes>>.
- [SHIEBER Stuart M.]. *"Don't ask, don't tell" rights retention for scholarly articles* [online blog] 18 Jun 2009 [Cited 2009-06-20]. Available from: <<http://blogs.law.harvard.edu/pamphlet/2009/06/18/dont-ask-dont-tell-rights-retention-for-scholarly-articles/>>.
- SPARC. *Author Rights: Using the SPARC Author Addendum to secure your rights as the author of a journal article* [online]. [Cited 2009-06-07]. Available from: <<http://www.arl.org/sparc/author/addendum.shtml>>. Also available in PDF format: <http://www.arl.org/sparc/bm~doc/SPARC_AuthorRights2006.pdf>.
- SPIVACK, Nova. *Understanding the Semantic Web and Twine Talk* [online]. [Cited 2009-06-07]. Available from: <<http://www.slideshare.net/novaspivack/nova-spivack-understanding-the-semantic-web-and-twine-talk>>.
- SWAN, Alma and BROWN, Sheridan. *Open access self-archiving: An author study* [online] 2005 [cited 2009-06-07]. Available from: <<http://eprints.ecs.soton.ac.uk/10999/>>.
- Together Again: Springer, Max Planck Agree To New "Experimental" Deal. *Library Journal Academic Newswire* [online]. 5 Feb 2008 [cited 2009-06-07]. Available from: <<http://www.libraryjournal.com/info/CA6528977.html>>.
- VERNOOY-GERRITSEN, Marjan, PRONK, Gera and GRAAF, Maurits van der. Three Perspectives on the Evolving Infrastructure of Institutional Research Repositories in Europe. *Ariadne* [online]. Issue 59, 30 April 2009 [cited 2009-06-07]. Available from: <<http://www.ariadne.ac.uk/issue59/vernooy-gerritsen-et-al/>>.
- Website resources**
- BASE. Bielefeld Academic Search Engine [online]. [Cited 2009-06-07]. Available from: <<http://base.uni-bielefeld.de/en/index.php>>.
- DOAJ. Directory of Open Access Journals [online]. [Cited 2009-06-07]. Available from: <<http://www.doaj.org/>>.
- DRIVER (Digital Repository Infrastructure Vision for European Research) [online]. [Cited 2009-06-07]. Available from: <<http://www.driver-community.eu/>>.
- DRIVER Mentor Service [online tutorial]. [Cited 2009-06-07]. Available from: <http://www.nottingham.ac.uk/toolkits/play_470>.
- DRIVER Studies [online]. [Cited 2009-06-07]. Available from: <<http://validator.driver.research-infrastructures.eu/validatorWeb/studies.do?action=load>>.
- DRIVER Validation [online]. [Cited 2009-06-07]. Available from: <<http://validator.driver.research-infrastructures.eu/validatorWeb/pages/validatorHome.jsp>>.
- DRIVER Wiki [online]. [Cited 2009-06-07]. Available from: <<http://www.driver-support.eu/pmwiki/>>.
- DRIVER. Documentation and Downloads. D-NET v1.0. [online]. [Cited 2009-06-07]. Available from: <http://www.driver-repository.eu/D-NET_release>.
- DRIVER. *Exposing textual resources with OAI-PMH : Guidelines for content providers* [online]. [Cited 2009-06-07]. Available from: <http://www.driver-support.eu/documents/DRIVER_guidelines_10.pdf>.
- DRIVER. Mentor Service. [online]. [Cited 2009-06-07]. Available from: <<http://www.driver-support.eu/mentor.html>>.
- DRIVER. Repository registration [online]. [Cited 2009-06-07]. Available from: <<http://validator.driver.research-infrastructures.eu/validatorWeb/pages/registrationInfo.jsp>>.
- eIFL-OA (Electronic Information for Libraries – Open Access) [online]. [Cited 2009-06-07]. Available from: <<http://www.eifl.net/cps/sections/services/eifl-oa>>.
- EPrints. *Journal Policies - Summary Statistics So Far* [online]. [Cited 2009-06-07]. Available from: <<http://romeo.eprints.org/stats.php>>.
- Europeana [online]. [Cited 2009-06-07]. Available from: <<http://www.europeana.eu/portal/>>.
- Google Scholar [online]. [Cited 2009-06-07]. Available from: <<http://scholar.google.com/>>.
- JISC-SURF. *Copyright toolbox* [online]. [Cited 2009-06-07]. Available from: <<http://copyrighttoolbox.surf.nl/copyrighttoolbox/>>.
- OALster [online]. [Cited 2009-06-07]. Available from: <<http://www.oalster.org/>>.
- OASIS (Open Access Scholarly Information Sourcebook) [online]. [Cited 2009-06-07]. Available from: <<http://www.openoasis.org/>>.
- OASPA. Open Access Scholarly Publishers Association [online]. [Cited 2009-06-07]. Available from: <<http://www.oaspa.org/>>.
- PKP. (Public Knowledge Project) [online]. [Cited Jun 7, 2009]. Available from: <<http://pkp.sfu.ca/>>.
- ROAR (Registry of Open Access Repositories) [online]. [Cited 2009-06-07]. Available from: <<http://roar.eprints.org/>>.
- Shambles. Plagiarism tools [online]. [Cited 2009-06-07]. Available from: <<http://www.shambles.net/pages/staff/ptools/>>.
- Sherpa/JULIET [online]. [Cited 2009-06-07]. Available from: <<http://www.sherpa.ac.uk/juliet/>>.
- Sherpa/RoMEO [online]. [Cited 2009-06-07]. Available from: <<http://www.sherpa.ac.uk/romeo/>>.
- Springer. *Springer Open Choice*. [online]. [Cited 2009-06-07]. Available from: <<http://www.springer.com/open+access/open+choice?SGWID=0-40359-0-0-0>>.
- Wageningen Yield (WaY) [online]. [Cited 2009-06-07]. Available from: <<http://library.wur.nl/way/>>.

Open Access and Repositories

A Status Report from the World of High-Energy Physics

Jens Vigen

jens.vigen@cern.ch

European Organization for Nuclear Research, CERN, Switzerland

Abstract: Access to previous results and their reuse in new research are at the very basis of scientific progress. In the era of e-Science and when the Open Access paradigm is changing scholarly communication, there is an unprecedented need for rapid and effective online access to scientific information. High-Energy Physics (HEP) pioneered innovation in scholarly communication with the invention of the Web, originally a vehicle of scientific information, and with the inception of online preprint repositories, introducing Open Access to preliminary scientific results. With the imminent start-up of the CERN LHC accelerator, one of the flagships of European science, the HEP community urgently needs a new platform for scientific information. Four international physics laboratories, in close collaboration with its partners in the publishing industry, have developed a vision to build such an innovative e-infrastructure: Inspire. The system will integrate present European and American databases and repositories to host the entire corpus of the HEP literature and become the reference scientific information platform of HEP worldwide. It will empower scientists with new tools to discover and access the results most relevant to their research; enable novel text- and data-mining applications; deploy new metrics to assess the impact of articles and authors. In addition, it will introduce the Web2.0 paradigm of user-enriched content in the domain of sciences. Inspire will be run on Invenio, an open-source platform that is scalable and portable to other fields of science.

In parallel the community is pushing for a complete change in the publishing model towards a scenario where the entire literature will be made available without subscription barriers. The project is run by an emerging consortium of HEP funding agencies, laboratories and libraries: SCOAP3 (Sponsoring Consortium for Open Access Publishing in Particle Physics). SCOAP3 will engage with scientific publishers towards building a sustainable model for Open Access publishing, which is as transparent as possible for HEP authors. The paper includes an outline of the history of Open Access in HEP, the details of the SCOAP3 model and the outlook for its implementation.

Keywords: open access; institutional repositories; digital publishing; High-Energy Physics

A short description of High-Energy Physics

The scientific goals of HEP are to unveil the intimate constituents of matter and probe their interactions. This is a quest as old as science, which today aims to attain a fundamental description of the laws of physics and the evolution of the universe, to explain the origin of mass and to understand the dark matter in the universe. HEP is an experimental and a theoretical science, with a community, counting some 20'000 members, split in roughly two halves: experimental and theoretical physicists. These scholars publish yearly about 6'000 articles. Of these, about 80% are articles produced by theoretical physicists and 20% by large collaborations of experimental physicists.

Experimental HEP scientists team in thousand-strong collaborations to build large scientific instruments, aiming to reproduce the energy densities of the universe at its birth. At the same time, theoretical particle physicists are linked in global networks through which they collaborate to formulate hypotheses and theories aimed to predict and interpret experimental findings.

HEP experimental research takes place mainly in international accelerator research centres while HEP

theoretical research takes place in hundreds of universities and institutes worldwide. However, these institutes also host experimental teams building parts of the large detectors used at the large accelerator laboratories and analyzing the data these collect.

CERN's Large Hadron Collider (LHC), the most powerful particle accelerator ever constructed, will start accelerating particles in 2009, after more than a decade of construction. The LHC program is at the technological frontier, and has required the invention, design and deployment of tools in engineering and information technology that did not exist at the time of the proposal of the scientific goals of the project. By now the first physics results from this endeavor should not be far away; consequently, publishers across the world are competing to be selected to publish the results.

Repositories

Scientific information in HEP has been at the IT frontier for many years. The arXiv.org repository, today owned and operated by Cornell University, is the lifeblood of our field. It went online in 1991 – this was even before the Web was born! The development of the Web led HEP to develop the first online databases of pre-prints, articles and books. However, opportunities offered in our private life, such as typing a snippet of a lyric in a search window and hearing a song played in a few seconds are still impossible in our professional life, where researchers cannot find a relevant article by just typing a few selected phrases. An example for all of a “dream service” would be feeding a figure into a search engine to identify the article to which it belongs.

Several HEP scientific information systems exist at the present. Currently the Deutsches Elektronen-Synchrotron (DESY), the European Organization for Nuclear Research (CERN), the Fermi National Accelerator Laboratory and the SLAC National Accelerator Laboratory build a new system, Inspire¹, as a collaborative effort. The vision is the construction of a single super-system for scientific information in HEP which will address all present needs of the community, as offered by the present systems. This super-system is a typical example of an e-infrastructure to enable e-science, which answers to existing precise needs of the community. It will serve as:

- the repository where all HEP Open Access articles will be hosted;
- a database which will offer new full-text and data-mining applications;
- a system to continuously measure the scientific production of individual countries and institutions, which is at the basis of cost sharing in Open Access publishing;
- a comprehensive, freely-available, citation index for HEP publications.

The system will provide the entire HEP community with services for e-science going beyond what is offered by today information systems:

- automatic selection of all articles of interest to the viewer of a given article, through a combined study of what previous users have read, citation analysis, author networks;
- automatic detection of the subject of an article from full-text mining and citation analysis;
- automatic choice of peer-reviewers from citation analysis and co-authorship patterns;
- a system of citation metrics aimed to tag influential, prestigious and popular articles, in order to complement

¹ Inspire [online]. [Cited 2009-07-07]. Available from: <<http://www.projecthepinpire.net/>>.

- the journal “impact factor” being now used to evaluate the productivity of institutions and individuals according to the journals in which they have published;
- access to numeric data related to figures.

Inspire will combine content from all existing relevant system and will be fully integrated with information providers, e.g. arXiv and the publisher portals, for the ingestion of new content. The system is being built on the technical platform Invenio², a software available under the GNU General Public License.

The current role of journals in the field of HEP

For over a decade between 90% and 100% of all HEP articles have, prior to publication, been provided by arXiv. In a recent paper Gentil-Beccot, Mele and Brooks³ illustrate this by plotting the arXiv-coverage of the content of the main peer-reviewed HEP journals as a function of time (Figure 1). It is worth noting that many HEP scientists routinely upload to arXiv a revised version of their preprint that matches the final peer-reviewed version, namely including any corrections introduced during the publication process.

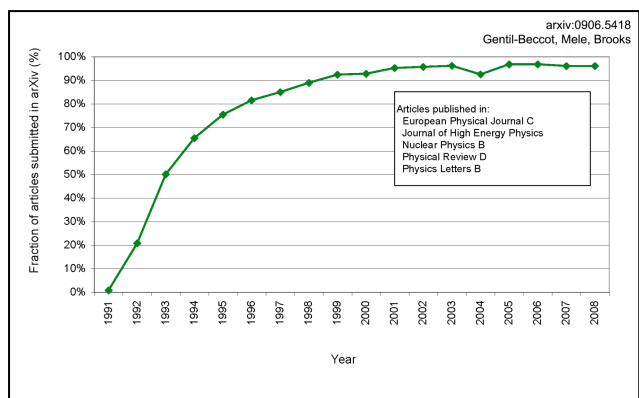


Figure 1 — Fraction of articles published in the main peer-reviewed HEP journals that also appeared, in some version, on arXiv.org as a function of time.

Further Gentil-Beccot, Mele and Brooks studied the reading behavior of the community by collecting the click streams generated by the SPIRES database, a community operated information service preferred by some 50% of the community for carrying out bibliographic searches.⁴ The outcome of the study, which might be considered surprising, is that less than 20% of the scientists actually

² Invenio [digital library software] Available from CDS [online]. [Cited 2009-07-07]. <<http://cdsware.cern.ch/>>.

³ GENTIL-BECCOT A., MELE S. and BROOKS T.C. Citing and reading behaviours in High-Energy Physics.

⁴ GENTIL-BECCOT A. et al. Information resources in High-Energy Physics.

opt for the published version, hosted on the publishers web site, when links both to the preprint and the article are made available side by side. The study does not contain data on why arXiv enjoys so much more traffic than the publishers, but it is likely to believe that the arXiv version is preferred by the community because it is freely available, carries the same information as the published version and it is only one click away while the publisher version is limited to subscribers and has to be picked up from an intermediate “splash page” that requires at least two clicks starting from SPIRES. Another surprising fact: only a very tiny fraction, less than 1%, of the scientists, uses the publishers’ portals to access articles.⁵ Obviously, the community-operated services are fully dominating the field; those who are not using SPIRES for searching are mainly using arXiv directly. So by combining the two pieces of information, Gentil-Beccot et al. estimate the advantage of arXiv over the published version to be around a factor eight.

This leaves the community with a conundrum; nobody reads journals, but everybody continue to submit their manuscripts for publication. The explanation for this is that arXiv is perfect for what concerns dissemination and access, but not for evaluation. The scientists therefore submit their works to journals to get the papers peer reviewed. Peer review remains, for any branch of science, an essential part of the scientific process and one of the main input to the mechanisms that regulate academic advancements.

Expressed in bullet points the role of journals, within HEP, is the following:

- Scientists do not read journals; they read arXiv
- Journals are for peer-review and officialdom
- Libraries’ subscriptions implicitly support this rather than buying access
- Growth of self-archiving is accelerating change
- We now see a convergence of repositories, peer-review, OA

Open Access

The goal of “Open Access” (OA) is to grant anyone, anywhere and anytime, free access to the results of scientific research. The OA debate has gained considerable momentum in recent years. It is driven mostly by two factors:

- The “serials crisis” of ever-rising costs of journals, which has forced libraries to cancel a steadily-increasing number of subscriptions, curtailing the access of researchers to important scientific literature.
- The increasing awareness that results of publicly-funded research should be made generally available. This need is amplified by the transformation of research activities

towards “e-Science”, carried out by a global scientific community linked by strong networks.

In December 2005 a tripartite Task Force, comprising funding agencies, publishers and research organizations, was set up to study the possibilities for OA publishing in High Energy Physics (HEP). Its main conclusion was that a model whereby the costs of publishing were paid globally rather than on an article-by-article basis was the most appropriate for a new publishing model. It is interesting to observe that in the past years all physics publishers have introduced Open Access options of one kind or another, clearly all moves that can be directly linked to the ongoing vivid debate in the community itself. However, in spite of that 90% of the articles published today in principle could have been published as OA, only a very small fraction of authors, due to the fact that no funding mechanisms so far are put in place to cover the corresponding publication fees, actually have the opportunity to choose this option. On the other side, there is no doubt that the community is motivated to communicate its scientific findings via OA; actually HEP pioneered OA way before the Internet facilitated information exchange as we know it today. For decades physicists actually shipped manuscripts intended for publication in 100s of copies, in the jargon referred to as preprints, to colleagues around the world for comments prior to publication. Librarians at CERN then came up with the idea of compiling a catalogue of these documents; a collection of papers that after all proved itself to be the backbone of communication within the community. This catalogue turned later into a database and eventually into a so-called “repository” containing collections of “pre-prints” freely accessible on the Internet. Today the main bulk of HEP pre-prints are available through repositories and the papers can easily be retrieved via services offered by arXiv, the CERN Document Server or SPIRES. Thanks to the speed with which they make results available, repositories have reinforced the role of preprints. This is what the OA movement often refers to as green OA. However, repositories do not perform peer review and may contain only the original versions of articles *submitted* to journals, and not necessarily the final, peer-reviewed, *published* versions.

Notwithstanding the success of repositories, there is consensus in the scientific community about the need for high-quality journals that will continue to provide:

- quality control through the peer review process;
- a platform for the evaluation and career evolution of scientists;
- a measure of the quality and productivity of research groups and institutes.

Making this class of papers available to anyone, anywhere and anytime brings us to what is referred to as gold OA.

The price of an electronic journal is mainly driven by the costs of running the peer-review system and editorial

⁵ GENTIL-BECCOT A. et al. *Information resources in High-Energy Physics*

processing. Most publishers quote a price in the range of 1'000–2'000 Euros per published article. On this basis we estimate that the annual budget for the transition of HEP publishing to OA would amount to a maximum of 10 Million Euros per year. In comparison, the annual list-price of a single “core” HEP journal today can be as high as 10'000 Euros; for 500 institutes worldwide actively involved in HEP, this represents an annual expenditure of 5 Million Euros.

The SCOAP3 model

The proposed initiative aims to convert high-quality HEP journals to OA, pursuing two goals:

- to provide open and unrestricted access to all HEP research literature in its final, peer-reviewed form;
- to contain the overall cost of journal publishing by increasing competition while assuring sustainability.

In this new model, the publishers' subscription income from multiple institutions is replaced by income from a single financial partner, the “Sponsoring Consortium for Open Access Publishing in Particle Physics” (SCOAP3). SCOAP3 is a global network of HEP funding agencies, research laboratories, and libraries. Each SCOAP3 partner will recover its contribution from the cancellation of its current journal subscriptions. This model avoids the obvious disadvantage of OA models in which authors are directly charged for the OA publication of their articles.

The financing and governance of SCOAP3 will follow as much as possible the example of large research collaborations and each country will contribute according to the number of its scientific publications, as presented in the appended figure. To cover publications from scientists from countries that cannot be reasonably expected to contribute to the consortium at this time, an allowance of not more than 10% of the SCOAP3 budget is foreseen.

In practice, the OA transition will be facilitated by the fact that the large majority of HEP articles are published in just six peer-reviewed journals from four publishers. Five of those six journals carry a majority of HEP content. These are *Physical Review D* (published by the American Physical Society), *Physics Letters B* and *Nuclear Physics B* (Elsevier), *Journal of High Energy Physics* (SISSA/IOP) and the *European Physical Journal C* (Springer). The aim of the SCOAP3 model is to assist publishers to convert these “core” HEP journals entirely to OA and it is expected that the vast majority of the SCOAP3 budget will be spent to achieve this target. The sixth journal, *Physical Review Letters* (American Physical Society), is a “broadband” journal that carries only a small fraction (10%) of HEP content; it is the aim of SCOAP3 to sponsor the conversion to OA of this journal fraction. The same approach can be extended to another “broadband” journal popular with HEP instrumentation articles: *Nuclear Instruments and Methods in Physics Research A* (Elsevier) with about 25% HEP content.

The schema will of course not be limited to the titles listed above; all publishers will be welcome to bid for being included. For new journals, criteria such as profile of the editorial board or the size of the author and reader base will be considered.

HEP has a natural overlap with related fields such as, but not limited to, astroparticle physics and nuclear physics. The five “core” journals include between 10% and 30% of articles in these disciplines, which will be naturally and logically included in the OA transition. This is in the interest of the readership and promotes the long-term goal of an extension of the SCOAP3 model to these related disciplines.

The fractions of “broadband” journals quoted above also include publications in these related disciplines.

Of course, the SCOAP3 model is open to any other, present or future, high-quality journals carrying HEP content. This will ensure a dynamic market with healthy competition and a broader choice.

The annual budget for the SCOAP3 operation will be established through a tendering procedure. The tender and the subsequent contracts with publishers will address the use of OA articles, the conditions for un-bundling OA journals from existing subscription packages, and the reduction of subscription prices for “broadband” journals following the conversion of a fraction of articles to OA. Provided that the SCOAP3 funding partners are ready to engage in long-term commitments, many publishers are expected to be ready to enter into negotiations along the lines proposed here.

Leading funding agencies and library consortia are currently signing Expression of Interest for the financial backing of the consortium (*Figure 2*). Once sufficient momentum is gained, the tendering procedure will take place determining the exact budget envelope. A Memorandum of Understanding detailing the financial contribution of each country and the governance of SCOAP3 will then be signed. Contracts will then be established with publishers in order to make Open Access publishing in High Energy Physics a reality.

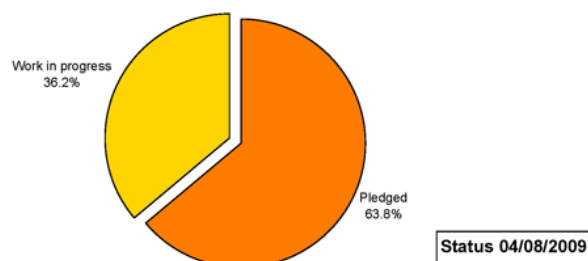


Figure 2 — Close to 2/3 of the estimated budget has been raised in only two years.

Outlook

The example of SCOAP³ will be an important milestone in the history of scientific publishing. It could rapidly be followed by other disciplines and, in particular, by fields related to HEP such as nuclear physics or astroparticle physics.

By achieving Open Access to the entire corpus of the literature in a given field and making this information available via the specialized repositories that have been developed over the last decade, a series of new opportunities will appear. Open Access is not eScience, but eScience will certainly require Open Access.

References

- GENTIL-BECCOT, Anne, MELE, Salvatore and BROOKS, Travis C. *Citing and reading behaviours in High-Energy Physics. How a community stopped worrying about journals and learned to love repositories* [online]. arXiv:0906.5418. CERN-OPEN-2009-007. SLAC-PUB-13693 [cited 2009-07-07]. Available from: <<http://arxiv.org/pdf/0906.5418v1>>.
- GENTIL-BECCOT, Anne et al. Information resources in High-Energy Physics : surveying the present landscape and charting the future course. *Journal of the American Society for Information Science and Technology* [online]. Vol. 60 (2009), no. 1, p. 150–160, arXiv: 0804.2701 [cited 2009-07-07]. Available from: <<http://www3.interscience.wiley.com/cgi-bin/fulltext/121383554/PDFSTART>> or from arXiv: <<http://arxiv.org/pdf/0804.2701>>.

Open Access and Institutional Repositories

from Local Initiatives to Global Solutions

Friedrich Summann

friedrich.summann@uni-bielefeld.de

Bielefeld University Library, Germany

Abstract: Bielefeld University and its library have a strong history of activities dealing with open access and services based on it. In 2003 the library introduced a OAI-compliant repository platform, mainly fed with electronic dissertations. In 2005 the university signed a resolution for supporting open access publishing for the scientific output. It was the task of the library to transfer this strategy into practice and so a repository for the peer-reviewed publications (postprints) was implemented called BiPrints. Recent activities have established an university publication platform called BieColl presenting local conference proceedings and a electronic journal. Current developments are focused on the integration of the university research information database and (funded in a national project) the development of the future version of the OPUS repository platform. In addition we started a funded project called Publister to support the integration of publication lists of scientists into the publication process. Recently we received the funding of project to combine the repository metadata information with automatically extracted classification annotations. All these activities aim for improving and optimizing the visibility and data quality of the scientific output of the university as a service of the library.

In 2001 the library started enhancing its technology with search engine technology as follow-up to earlier metasearch developments we did before. While in the beginning the implementation focussed on crawling web-sites (as search engine do mainly) the work has been shifted to processing OAI-PMH harvested metadata and led to the Bielefeld Academic Search Engine (BASE) which now is a leading application among the non-profit OAI service providers. Meanwhile it has integrated around 1250 repositories world-wide and realised a list of integrated services as linguistic features, classification browsing and end-user functionality (drill-down, search history, sorting, country search).

The combination of a strong role in repository strategy and the BASE expertise were the reasons why Bielefeld UL became a participant of the European Union funded DRIVER project (Digital Repository Infrastructure Vision for European Research) aiming to support the network of scientific repositories in Europe. Our efforts were related to the technical topics Harvesting, Aggregating and Indexing while in parallel we worked for improving the network infrastructure.

Keywords: institutional repositories, open access, search engine

Bielefeld University and its library have a strong history of activities associated with open access publishing and services based on or related to it. Meanwhile this includes broad initiatives and projects on the local, national and international level.

The local approach

Bielefeld University Library was involved in a state project to establish a so-called state-wide digital library in the late nineties. This system (Digital Library Northrhine-Westphalia) included the storing and presentation of all electronic material available (early examples were local digitisations and dissertations). While in this first phase those electronic resources have been stored in the file system and have been presented using static html pages the need for a comfortable system supporting the integrational workflow and including a functional end-user interface became very clear. So we decided to use the PHP based repository system OPUS (developed by Stuttgart University

Library) and established BieSON¹ (Bielefeld Server for Online Publications). This OAI-compliant repository included mainly dissertations and few additional material (non peer-reviewed) and started in 2004. After the Berlin declaration Open Access became a major theme in Germany and so Bielefeld University signed a resolution to support open access publishing in June 2005. As a consequence the library started BiPrints² as the repository system for the scientific output (peer-reviewed publications) of Bielefeld University. BiPrints and another platform called BieColl³ for electronic publishing (conference proceedings and a journal archive) are based on OPUS as well but a lot of efforts had to be done to add advanced functionalities related to the specific adoption. Besides that we introduced a virtual repository BIETAS⁴ to

¹ BieSON <<http://bieson.uni-bielefeld.de/>>.

² BiPrints <<http://repositories.uni-bielefeld.de/biprints/>>.

³ BieColl <<http://biell.uni-bielefeld.de/biell/>>.

⁴ Bielefeld University Library. Digital full-texts <<http://www.uni-bielefeld.de/english/diglib/>>.

make all distributed material (digitizations) OAI visible. All these repository platforms have been registered as official OAI data providers. Obviously we have paid attention to let them listed in the relevant registries (OpenDOAR, ROAR etc.) as well.

Bielefeld University is a member of DINI (German Initiative for Network Information) to enforce the local strategic position within the German open access activities since 2005. The BieSON repository owns the DINI certificate as approved installation.

To support publishing of Bielefeld authors the library signed a contract with BioMedCentral and supported the publishing of a journal at the state-wide platform DiPP⁵ (Digital Peer Publishing). To respond to the demand of electronic publishing of local journals the library experimented with a test installation of OJS (Open Journal Systems) and the first publication will be activated soon.

The role of the library within the university and among the other IT-based departments is a national discussion topic since about 2000. There were a lot of different approaches to solve the uprising problem of concurrent activities and some national tenders to support the process of finding solutions. The strategy followed in Bielefeld was a dual-head solution with a CIO for Information technology (with a Computing Centre background) and a CIO for academic publication located in the library. Thus the essential library IT activities include all publication and database related applications. The overall goal of this efforts was to bundle and foster the publishing activities and to enforce the visibility of the local academic output. Today the library IT is independent and has undertaken some general activities as the university e-learning system and all search and publication related applications.

A specific problem in this context is the coordination of university research projects. CRIS systems (current research information systems) to organize and document the complete project activities and output do not play a strong role in Germany. This is the situation in Bielefeld as well and we have an application called Forschungsdatenbank⁶ (Research Database) which displays the project-related information including the publication list. A current local activity is to coordinate the uploading of project publications between research database and local repository.

The national approach

The role of Bielefeld University Library as an innovative future-orientated institution has led to a lot of state and federal funded projects in the past. In the OA context the local activities have prepared the path for participations in national-funded projects. There came up some national calls for tenders in the open access context in the last years in Germany and the library backed by its

earlier activities has prepared a number of proposals, mostly together with partners.

The Open-access.net project⁷ (financed by the DFG, German Science Foundation) realises an information platform to bring together all information on the subject of Open Access and to distribute it to the players in this field.

Together with 8 other partners (libraries and regional library unions running local or regional platforms) a consortium has been set up to improve the OPUS repository software which is broadly used in Germany in more than 60 repository environments. This piece of software built on PHP and MySQL is easy to use and has been adopted by both aggregated platforms and local installations.

Based on the adopting developments with OPUS Bielefeld UL became a partner of this DFG project. The outcome OPUS4 will deliver an upgraded modern version of the repository system in summer 2009. The Bielefeld work packages deal with collection support and the interoperability with research information systems and the DRIVER guidelines.

In April 2009 the DFG project PubLister started with the focus on integrating scientists' publication lists into the academic publishing process including the repository workflow to collect the necessary metadata. This approach tries to make the publication and repository workflow more efficient. First experiences show that the current situation includes very different approaches and that it will be necessary to support the production of these lists to make them more inter-operable.

As the latest project in this field we just have received the funding for a DFG project for automatic extraction of classification information from scientific documents. In cooperation with the Computer Linguistic Departments of Bielefeld and Leipzig University their text mining technology will be used to build up a broad knowledge base from OAI metadata and the corresponding documents. This will serve as a linguistic instrument to process statistical comparing activities with fulltexts with as a result a new or refined DDC classification information. DINI-certificated repositories and all OPUS repositories include DDC information and this is the starting point for establishing the DDC classification browsing in BASE. Eprints uses LOC annotations per default and some efforts will be taken to include such information via mapping crosswalks. The project points on producing an open interface for interested partners which may call the engine with text material and will receive a response including DDC codes. BASE will serve as a showcase to test and integrate the interface and to evaluate the results.

The international approach

The scientific search engine BASE⁸ (Bielefeld Academic Search Engine) grew out of efforts to combine library-orientated search systems with the up-coming search

⁵ DiPP NRW <<http://www.dipp.nrw.de/>>.

⁶ Forschungsdatenbank <http://ekvv.uni-bielefeld.de/foda_publ/publ/Home.jsp>.

⁷ Open access <<http://www.open-access.net/>>.

⁸ BASE <<http://www.base-search.net/>>.

engine technology. In Germany it became clear that both students and scientists more and more used search engines in their daily work. Thus the idea was to combine the end-user comfort and functionality with the depth and data quality of library applications. Shortly after the start in 2002 Open Access and especially OAI-PMH became more and more successful and so it was obvious to use the OAI-PMH protocol to collect scientific metadata with a certain bibliographic quality. In the beginning phase we used the standard search engine approach to collect data by crawling web-sites and files delivered by partners. Both strategies faced problems, crawling the minor data quality and exported files the need for very specific crosswalk scripts. Both problems are not appearing when using the OAI-PMH interfaces with an increasing coverage of academic institutions. Today more than 98 % of the data are OAI-PMH responses.

A broad fundament of harvesting and aggregating expertise was developed during these years. Lots of pre-processing scripts have been written to handle the variety of data formats – related to the many existing interpretations of the protocol and simple faults in realizing it.

Technically BASE consists of two independent parts, backend and frontend. The backend as the indexing and query-processing part is using the commercial FAST search engine package (now Microsoft), originally developed as a Google competitor on the search engine market (alltheweb.com). The frontend realizing the end-user interface has been developed by Bielefeld University Library and the php scripts are using the standard FAST xml-http interface. This communication has been enhanced with a common interface which currently is in use in Germany by library portals (Digital Library North Rhine-Westphalia, Karlsruhe Virtual Catalogue) and a meta search engine (Metager). Locally we use this interface in our local library catalogue to offer additional references for electronic resources.

After a rapid growth BASE today covers more than 1250 scientific repositories from all around the world pointing to almost 20 million documents, mostly freely accessible. The BASE search environment offers a broad range of advanced search functions (truncation, drilldown, linguistic features, sorting, search history, classification browsing, Google Scholar integration). Recently a country search has been introduced which allows the end-user to search in national sets of repositories.

The current status of BASE is built on harvesting efforts in the last seven years (testing around 2000 OAI interfaces) which have generated a broad expertise on all aspects of this area – challenges, problems, behaviour of different platform systems (down to the version level) and a long list of mapping rules to improve the heterogenous content interpretation and metadata quality.

We started our efforts as local initiative and ended as an international academic search engine – reacting on developments in the academic search technology. Anyhow, the expertise of BASE and the strong role in the OA community were the reasons why Bielefeld UL became a

participant in the European DRIVER⁹ project (Digital Repository Infrastructure Vision for European Research) aiming to support the network of scientific repositories in Europe. The project started in June 2006 with the goal of establishing a testbed with integrating 51 repositories from 5 countries. Phase 2 (Dec. 2007 – Nov. 2009) had the goals of establishing a productive infrastructure, expanding the geographical coverage, extension of the metadata and data formats and the extension of services. Phase 1 included 10 partners from 8 countries (Greece, Belgium, Italy, France, Netherlands, Poland, Great Britain and Germany) while in Phase 2 three additional project partners from Portugal, Denmark and Slovenia and additional partners from Sweden, Norway, Finland, Lithuania, Spain and Ireland joined the consortium. The coordination is done by the Univ. of Athens and the State and University Library in Göttingen.

Bielefeld UL is a partner in the technical group which has designed and realized the open source software release for repository services. The design is strictly SOA-oriented and includes a long list of different services communicating via Web Services. We are dealing with the tasks Harvesting/Aggregating, Storing and Indexing. The technical team in Bielefeld (2 developers) has developed the SOA-compliant infrastructure for the corresponding modules.

Because of the open source background the commercial BASE background had to be put aside and we switched to a Lucene-based solution.

Though there is a relevant role for technical developments DRIVER is an infrastructure project in the first instance. Because of its practice-orientated background Bielefeld UL is playing a strong role in the repository network infrastructure activities, together with the more library-orientated partners from SURF (Netherlands), Univ. of Nottingham and Univ. of Ghent (Belgium). An essential activity was publishing the DRIVER Guidelines¹⁰ for repositories. This paper, now available as version 2.0, defines a list of rules and recommendations for the usage of OAI-PMH and Dublin Core to support the delivery of high-quality metadata in a stable and efficient way. Implementing the guidelines should improve the deployment of visible and stable repositories based on a high standard quality of metadata and services.

The DRIVER validator development was planned as additional instrument to proof the specific DRIVER guidelines compliance and to compute a score for the degree of DRIVER inter-operability. This tool is embedded in the DRIVER registration process started by the local repository manager and is integrated in the DRIVER Helpdesk support activities. The concrete work of harvesting and aggregating including the feedback communication with the repository managers is currently managed in Bielefeld. Currently the DRIVER information

⁹ DRIVER <<http://www.driver-repository.eu/>>.

¹⁰ DRIVER Guidelines 2.0 <http://www.driver-support.eu/documents/DRIVER_Guidelines_v2_Final_2008-11-13.pdf>.

space covers more than 200 institutional repositories from Europe, currently with a focus on the UK, Holland, Germany and Scandinavia and up-coming countries as Spain and Portugal.

The situation in Czechia and Slovakia is similar to most of the Eastern Europe countries and therefore it is a DRIVER goal to support the repository infrastructure. BASE has integrated 2 repositories from Czechia (Technical Univ. Ostrava, National Academy of Science) and none from Slovakia, Within DRIVER we started the integration of the University of Pardubice DSpace system which has been registered shortly.

Besides that we are participating in the European PEER¹¹ project (Partnership for European Environmental Research) with the focus on the relationship between publishers, repositories and the research community. The project will evaluate the bi-directional influence of the different strategies of publications by comparing access statistics of articles on publishers web-sites with or without parallel open access availability via an institutional repository. More than 50,000 European accepted manuscripts from up to 300 journals will become available for archiving within repositories. The Bielefeld repository will be one of the test platforms.

Conclusion

Bielefeld University Library has a reputation as an innovative and pragmatic institution.

The listed activities and projects with focus on open access publishing and repository systems have a strong coherence in practice and in fact it makes no difference if

the project resides on the local, national or international level. Expertise and development results have to be used vice-versa and so communication between the projects and open interfaces is essential. The integration of applications and thus the development and usage of interfaces has always been a major focus of the libraries IT efforts. Many of the results of this bundle of activities will have practical values for others and it will be our task to enforce these benefits for the university staff and the end-users of our services.

Website resources

- BASE (Bielefeld Academic Search Engine) [online]. [Cited 2009-06-07]. Available from: <<http://www.base-search.net/>>.
- BieColl (Bielefeld eCollections) [online]. [Cited 2009-06-07]. Available from: <<http://biecoll.ub.uni-bielefeld.de/biecoll/>>.
- Bielefeld University Library. Digital full-texts [online]. [Cited 2009-06-07]. Available from: <<http://www.ub.uni-bielefeld.de/english/diglib/>>.
- BieSO (Bielefelder Server für Online-Publikationen) [online]. [Cited 2009-06-07]. Available from: <<http://bieson.ub.uni-bielefeld.de/>>.
- BiPrints [online]. [Cited 2009-06-07]. Available from: <<http://repositories.ub.uni-bielefeld.de/biprints/>>.
- DiPP NRW (Digital Peer Publishing NRW) [online]. [Cited 2009-06-07]. Available from: <<http://www.dipp.nrw.de/>>.
- DRIVER (Digital Repository Infrastructure Vision for European Research) [online]. [Cited 2009-06-07]. Available from: <<http://www.driver-repository.eu/>>.
- DRIVER. *DRIVER Guidelines 2.0 : Guidelines for content providers - Exposing textual resources with OAI-PMH* [online]. November 2008 [Cited 2009-06-07]. Available from: <http://www.driver-support.eu/documents/DRIVER_Guidelines_v2_Final_2008-11-13.pdf>.
- Forschungsdatenbank [online]. [Cited 2009-06-07]. Available from: <http://ekv.uni-bielefeld.de/foda_publ/publ/Home.jsp>.
- Open access : Open access to scholarly information. [online]. [Cited 2009-06-07]. Available from: <<http://www.open-access.net/>>.
- PEER (Partnership for European Environmental Research) . [online]. [Cited 2009-06-07]. Available from: <<http://peer-initiative.org/html/>>.

¹¹ PEER <<http://peer-initiative.org/html/>>.

Users' Behavior and Institutional Repositories

Towards Information Ecology

Jela Steinerová

jela.steinerova@fphil.uniba.sk

Comenius University Bratislava, Faculty of Philosophy, Slovakia

Abstract: Users' information behavior in the electronic environment is analyzed with respect to pragmatic information use of the Net generation. Results of two studies on human information behavior in libraries and relevance assessment in Slovakia are briefly reported. Academic information electronic environment is determined in context of building university cyberinfrastructure and institutional repositories. Pragmatic information style prevails in users' behavior with typical preferences of easy access, social networking, immediacy and visualization. Changes in relevance judgments in the electronic environment are determined including rich context and creative discovery. The concept of information ecology is briefly outlined as an emergent topical issue of information science based on knowledge of information behavior in the electronic environment. Information ecology and institutional repositories are discussed with the emphasis on ecological reuse of information objects and tools for knowledge organization. Ecological characteristics of academic information environment are determined as mixture of sources, systems and services managed by a social actor. A model for research into information ecology of the academic information environment is introduced. The proposed research includes deeper surveys of users and experiments with open source tools for organization of information. The contribution of information ecology concept to practice of systems, services and products of memory institutions is summarized.

Keywords: academic information environment; users' information behavior; institutional repositories; information ecology model

Introduction

Success of libraries has always been determined by focus on users. Based on new features of electronic environment information use has been subject to constant adaptation and innovation. Until recently, libraries operated on a paradigm of stable, structured and verified sources of information. New web x.0 environments evoke new types of information behaviors based on the use of new media. A new paradigm can be characterized by changing products and services as response to changing information needs. One predetermined structure does not fit all. Information, services, products and users are in flux.

Human information behavior in the electronic environment is marked by easy access, quick online reading, and social networking. Cognitive mapping and visualization are important for reduction of cognitive load in information processing and use. Practices of collaborative information use in web 2.0 change the traditional library paradigm. The balance between tradition and modernity of the electronic world is manifested by changes in users' information behavior and new tools and services of digital libraries.

In this paper we report on results of two projects aimed at library users' behavior in Slovakia and on relevance judgments of doctoral students. Research results are compared to similar user studies of major library and information research centers (OCLC, CIBER). Based on results and analyses we derive a new concept of information ecology of the academic information environment including experimental use of open access tools in university repositories.

Users' Information Behavior in the Electronic Environment

Many authors and surveys noticed challenge hidden in new patterns of information use by young people called the "Net Generation". Their habits caused emergence of new policies and practices in higher education and deep, revolutionary changes in academic information environment. Young students are digitally literate, prefer visual information processing, inductive discovery, fast response time.¹ The Net Gen exhibits tendency to socialize and participate in networks. Although many surveys generalize their patterns of information use, the information strategies of individuals may be unique.

For example the *The Valley of the Shadow* archive² shows how students learn by experience in electronic environment. New types of „first-person learning“ by doing emerge. Interaction, simulation and visualization play an important part and experiential learning is similar to computer games (e.g. Digital Based Game Learning). Young people are impatient, they prefer scanning to reading, they are „bouncers“, „fast surfers“ etc. Some authors³ say that movements in the electronic environment change mental models and thought of this generation. It is marked by image and spatial representations, non-linearity, parallel processing, immediacy.

¹ OBLINGER D.G. and OBLINGER J.L. (eds.) *Educating the Net Generation*.

² *The Valley of the Shadow* <<http://www.iath.virginia.edu/vshadow2/>>.

³ PRENSKY M. *Digital Natives, Digital Immigrants*.

For university managers, teachers and academic librarians this means that the whole traditional paradigm of education and learning should be transformed. It is important to understand students – what is important to them and how they learn best and work with information. They would not read extensive texts, they are experiential learners. That is why universities invest into IT infrastructures, e-learning and university repositories. However, in social sciences and humanities the traditions of critical reading still persist. The challenge is how to accommodate prevailing information and learning styles of the young Net Generation with needs of building knowledge from pathways through information space. Accessibility and ownership of information are sometimes interpreted as goal, but information use is based on a different paradigm – on transformation of information into knowledge.

Information Styles in Academic Information Environment

We regard information style as generalization of typical ways and preferences regarding access and information processing. Academic information environment can be determined as information, human, technological and organizational resources aimed at support of information activities in learning, scholarship and academic management. Our studies into users' information behavior of academic libraries in the project on Interaction of Man and Information Environment confirmed two information styles of users' behavior, i.e. the pragmatic and analytic styles. The pragmatic style prefers simple access to information, simple organization of knowledge, low cost and fast access to electronic resources (quantity and time). The analytic style is marked by deeper intellectual information processing. It emphasizes reliability and verification of information and sources and reviewing process (quality and relevance). In our surveys we found out that pragmatic style of information processing dominates. For surveys we developed an original methodology which integrated knowledge from social and cognitive science into information behavior.

	pragmatic	analytic
seeking	horizontal	explorative
terminology	clear, simple	multidisciplinary
assessment	surface, serendipity	experience in relevance judgments
organization	surface, field dependence	integrative, based on expert knowledge and experience
planning	intuitive, simple queries	complex queries
purpose	orientation	intellectual processing
emotions	trust, optimism	doubts
motivation	fast solution	understand contexts
access	navigation	interpretation

Table 1 – Differences between pragmatic and analytic styles.

Typical differences between these two styles are summarized in *table 1*.

Based on these results we developed models of information behavior and recommendations for information literacy.⁴ We suppose that further surveys of information behavior styles can help form information strategies and information ecology of academic information environment.

These results correspond with similar studies⁵ which determine patterns of information behavior of students and academics in the electronic environment. Simple access and „Google“ paradigm of information seeking and use are preferred. Even the use of time and space is changing.⁶ The question is how academic libraries and information strategies of universities can respond to personalization and socialization in learning spaces. On the other hand, questions of provenance of digital collection and cognitive authority are appreciated in conduct of scholarly research. Ownership and control of texts in institutional repositories also influence information styles and strategies of information behavior in the academic environment.

Prevailing pragmatic information behavior styles distorted traditional formal communication in higher education and scholarship including traditional academic librarianship. Academic libraries and universities started building large digital repositories. The problem is that students behave like digital consumers in general digital spaces and that social networking and consumer websites change their expectations as for the forms of education, learning and scholarship

Relevance Behavior of Doctoral Students

Another key part of information ecology is the concept of relevance. In the second project on Information use and information behavior we investigated relevance judgments of doctoral students.⁷ The study of relevance as part of information use by 21 doctoral students was designed within the phenomenographic concept. Semi-structured interviews as a main method of gathering data were applied. Students were selected from social sciences and information science at the Faculty of Philosophy, Comenius University Bratislava, Slovakia. Research questions concentrated on perception of relevance, manifestations of relevance in the electronic environment,

⁴ STEINEROVÁ J. *Informačné správanie*.

⁵ CONNAWAY S.L. *Make room for the Millennials*. PRABHA Ch. *at all. What is enough? Information Behaviour of the Researcher of the Future. Sharing, Privacy and Trust in Our Networked World*. SCHMITZ D. *The Seamless Cyberinfrastructure*.

⁶ DEMPSEY L. *Always on*.

⁷ STEINEROVÁ J., GREŠKOVÁ M., ŠUŠOL J. *Prieskum relevancie informácií*.

ways of categorization and typology of relevant information. Results were visualized in concept models.

Respondents confirmed relevance as value, utility and importance. Main findings suggest that the same criteria are used through different contexts and are related to development of information needs. Findings suggest that relevance judgment is multidimensional, based on multi-criteria cognitive processing. Relevance is experienced and integrated by emotions, especially delight, discovery and anger. The findings have implications for design of library and information systems and services and for information literacy. Users need support for discovering, decision-making and participation. Policy makers should consider relevance behavior in information strategies of educational and information institutions.

	Traditional library environment	Electronic environment (internet, web, digital library)
System relevance	Bibliographic level, formal match	Interactivity, complexity of relations
Topical relevance	Orientation, navigation to sources	Content, intellectual - knowledge
Criteria	One-dimension	Rich multi-criterial
Organization of Information	Linearity	Non-linearity
Added value		Context, visualization, collaboration
Cognitive relevance	Cognitive state, style, user categories	Personalization, intelligent interfaces
Information structures	Deterministic	Flexible, concept linking, Mosaic-like
Communicative relevance	Reference interview, communication styles	Group sorting, recommendation, dialogue
Situational relevance	Tasks, problems, situations	Uncertainty reduction by contexts, links, organization
Motivational relevance	Preferences, goals	Easy access, service forms, discovery, creativity

Table 2 — Differences in relevance judgments between library and electronic environments.

Concept maps can be used in new knowledge organization in digital repositories. New models of relevance in the electronic environment were derived which emphasize interaction, linking, visualization of concepts and relations, as well as on community collaboration. New relevance models emerge as opposed to traditional retrieval tools (e.g. Google). The most important is context, ranking and linking. These features are integrated into new relevance concepts in the electronic environment. In table 2 we depict the differences between relevance in traditional and electronic environments.

These differences show especially different contexts of users (who, where, when and for which purpose uses information). Connections between electronic sources in institutional repositories and users' relevance behavior lead

to new concepts of e-learning and digital scholarship. A possible way for interpretation and examination of these complex phenomena is information ecology.

Information ecology and information behavior

The concept of information ecology emerged from information management and information behavior studies. It is based on complex relationships between humans and technologies while using information in communities and organizations. Information ecology was determined by Davenport and Prusak⁸ as making information meaningful. Information ecology as a metaphor should help manage (academic) information environment, especially integrate diverse types of information and information behavior of people. Another concept of information ecologies⁹ is based on relationships between information technologies and people in transforming information to knowledge. Information ecologies represent procedures, goals, values of communities supported by technologies. Information ecologies are places where people use tools and in social relations help each other in information activities.

Based on environmental psychology an ecological constructionist model of user information behavior¹⁰ was developed which integrates affective, cognitive and sensorimotoric parts of information activities. Ecological model of information seeking and use¹¹ depicts a social actor involved in such settings as information needs, specific personal, physical, working and social contexts. The features of adapting and monitoring of information environment are here emphasized. Information ecology is also connected with studies of affective information behavior.¹² Emotions influence receipt and use of information as well as relationships to information technologies. D. Nahl explained the affective revolution in information science, especially the affective filter in information seeking and evaluation of information. Emotional entanglements in information behavior of students play an important part in the academic information environment. A model of the affective information behavior ecology¹³ proves macro- and micro-emotional contexts that shape students' information behaviors. The author calls for more attention to affective dimensions of services and products in the academic information environment.

⁸ DAVENPORT T.H. and PRUSAK L. *Information Ecology*.

⁹ NARDI B.A. and O'DAY V.L. *Information Ecologies*.

¹⁰ NAHL D. *A discourse analysis technique for charting the flow of micro-information behaviour*.

¹¹ WILLIAMSON K. *Ecological Theory of Human Information Behavior*.

¹² NAHL D. and BILAL D. (eds.) *Information and Emotion*.

¹³ GIVEN L.M. *Emotional Entanglements on the University Campus*.

Our concept of information ecology is based on common ideas of the authors who emphasize meaningful information activities in information environment. It is based on knowledge of information behavior of different social actors, but also on tools that can help clean information environment. In search of these tools we emphasize better knowledge organization in electronic environments and institutional repositories, information literacy and relevance judgments.

Institutional repositories and information ecology

Institutional repository can be understood not only as digital sources based on intellectual outputs of academic institutions, but also as a set of services, products and tools that document the intellectual and cultural life of a university or other institution.¹⁴ In fact, it is a special case of cyberinfrastructure marked by many different document forms, information objects and materials that support education and research. It is closely connected with academic community composed of social actors (producers, users, designers, managers, librarians, etc.). Some authors state that a repository is represented by digital library and a publishing system.¹⁵

The context of building institutional repositories of universities is represented by deep paradigmatic changes in digital scholarship and e-learning and e-research.¹⁶ It can also be framed within the contexts of building digital libraries and interactive intelligent environments that support intellectual activities in information production and use. Most universities in developed but also less developed countries started to build institutional repositories and digital archives of students' and faculty works. The international UNESCO Networked Digital Library of Theses and Dissertations (NDLTD) is noteworthy. However, the organization of knowledge based on concept mapping or other efficient semantic tools in these repositories is still a challenge for practice of information science.

Ecological aspects of these environments can be explained by three principles: 1. information services and systems are parts of the holistic complex information process of the institution and are similar to natural organisms, 2. human information behavior is ecological in that it links external and internal knowledge, emotions, and social networks, 3. in digital environments it is possible to manage the ecological reuse of information objects, services and products. The tools for satisficing and optimizing of information processes are integrated in specific knowledge organization tools in digital environ-

ments. As examples we can mention concepts maps, topic maps, intelligent thesauri or complex ontologies. We regard these tools as crucial in information ecology research as cleaning filters that support sense-making in the individual, group and organizational information environments. For experiments most common open software tools used are such as DSpace, e-Prints, Fedora, Greenstone.

User issues related to institutional repositories can be divided into implementers, depositors and end users. Several studies of end users of institutional repositories confirm that internet search engines, especially Google is most popular with undergraduates, while scholars and graduate students tend to use more specialized and professional digital resources and tools.¹⁷

The role of information professionals in academic information environments is then not only information processing and knowledge organization, but also support of natural information activities and pathways. This support is represented by navigation, filtering, and help with cognitive and affective overload in information behavior. Re-conceptualization of scholarly communication changes not only traditional information professions of librarians, publishers, scholars, teachers, but also institutions as libraries and universities. Open access movements introduced also the concept of open repositories. It means free accessibility of outputs of scholarly communication and democratic access to sources for education with rules for intellectual property rights and principles of knowledge commons. Repositories can be sources for collaboratories, virtual laboratories and socio-technological infrastructures. Good examples include Humane Genome, e-Horizon, European DRIVER, MIT courseware etc. Digital archives and repositories have already their history. The first was the famous arXiv.org in physical sciences. Other well-known repositories include CogPrints (psychology and cognitive sciences), PubMedCentral (medicine), or RePEc (Research Papers in Economics). Important differences can be noted between sciences, social sciences and humanities. From a number of initiatives in building repositories we have to mention the activity of Slovak academic libraries in the project ETD SK (Electronic Theses and Dissertations Slovakia, <<http://www.etsd.sk>>).

Information ecology in our concept draws also on the activity theory.¹⁸ The information activity can be modeled in the Engeström's model triangle composed of the three components: man, information objects and tools. This can be combined into such activities as production, consumption, change and distribution of information. Besides communicative tools an important role is played by community, values, division of labor, and collaboration. In the information process we emphasize the principles of cognition and communication, adaptation and interaction, relevance, satisficing and optimization.

¹⁴ SCHMITZ D. *The Seamless Cyberinfrastructure*.

¹⁵ BORGMAN Ch. *Scholarship in the Digital Age*.

¹⁶ BORGMAN Ch. *Scholarship in the Digital Age*.

LYNCH C. *The Institutional Challenges of Cyberinfrastructure and e-Research*.

Van de SOMPEL, H. et al. *Rethinking Scholarly Communication*.

¹⁷ SCHMITZ D. *The Seamless Cyberinfrastructure*.

¹⁸ WILSON T.D. *A re-examination of information seeking behaviour in the context of activity theory*.

Thus, information ecology can be determined as a way of information existence in the information environment. Within the goal of information activities it is not the volume of information or immediacy of delivery that matter but tools for support of making sense and transformation of information to knowledge.

Information ecology of the academic information environment

We suppose that proposals for protection of information environment at individual, group and institutional levels are needed. The following questions rise: How to organize digital information environments to make information use more efficient? How to provide contexts for sense making? Information ecology should identify factors that make an impact on the information environment and tools for eliminating information overload, redundancy and risks of information use. The components of information ecology are people, intelligent technologies, and tools for knowledge organization. Ethical, legal and security precautions while sharing information in digital repositories are parts of information ecology.

Transformation of academic libraries from passive repositories to adaptive systems with interactive organization of information based on information behavior is part of information ecology. Intelligent technologies and cyberinfrastructure can enhance ecological aspects of information use. Ubiquitous and unobtrusive technological mediation and knowledge of cognitive and affective principles of human information behavior can help form ecological models of information use.

Institutional repositories and digital libraries in ecological perspective include interactions with objects, sorting and selections. They support not only adaptations of humans to information environment, but also discovery of new information, experimental analyses, concept and topic mapping and data and text mining. Based on these ideas we developed a model for research of information ecology of academic information environment (*Figure 1*).

In the centre of the activity triangle we place the social actors (teachers, researchers, students, managers) who perform information activities and use tools for manipulating digital objects. The setting is the space of digital library and institutional repository. The model outlines two components: activities of social actors: production, registration, certification, distribution, archiving and updating, and components of digital repository (information objects, communicative activities, added value, representation, technologies). In interactions between these two components we can follow continuity and discontinuity of scholarly communication in digital environments (formal and informal communications, collaboration, quality control, information overload, knowledge production).

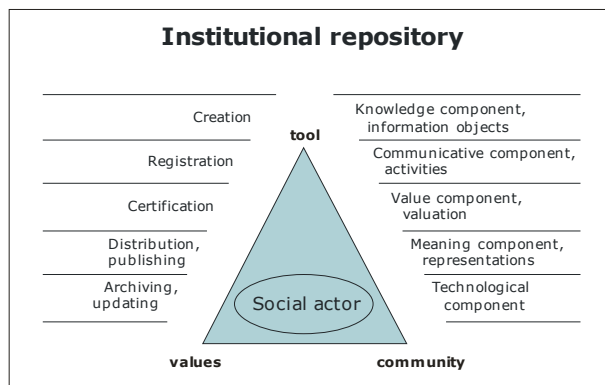


Figure 1 — Model for research of information ecology of the academic information environment.

Using quantitative and qualitative methods of research we plan to organize surveys of information behavior of social actors in the academic community with the emphasis on electronic environment and pathways of students, researchers, and managers. Important perspectives will include quality assessment, organization of information, publishing, collaboration, information overload. In the second part we plan to run experiments with organization of information sources in digital repository based on a sample of electronic students' works. We will compare several open source tools for knowledge organization (E-Prints, D-Space and C-maps). Concept maps could help transform information to knowledge in different functions of academic communication. Finally, proposals for support of efficient information behavior and methodologies for structuring information in university repositories will be elaborated.

Conclusions

Information ecology in our concept is based on human information behavior and information systems. Institutional repositories are parts of academic cyberinfrastructures and support data management, data curation, analyses, mining and visualization. The context of the proposed model is represented by changes in scholarly communication and education in digital environments. New knowledge organization tools can help manage information environment and information overload. The proposed model synthesizes our results of two research projects. We compared information styles and relevance judgments in traditional and electronic environments.

Information habits of students in the electronic environment indicate that simplicity, attractiveness and immediacy dominate the information use. In this respect new models of information ecology should put emphasis on navigation, personalization and visualization. New approaches to support of information literacy are required, especially in judging relevance, concept mapping and respect to copyright. Explosion of digital content has to be harmonized with organization and access to information.

Designers of systems and services can be inspired by reconstructed pathways of social actors in academic digital environments and by community social networks. In scholarly communication services should be tailored to groups and changing information needs.

Information ecology can help sense-making and information use in the academic information environment. While in technological concepts people had to adapt to technologies, in information ecology information processes and technologies are adapted to information needs and become parts of scholarly communication. The idea of "the networked social library" becomes true by interactive features of collaboration, creativity and community building. Institutional repositories should be open not only to information use, but also to production and social interaction. The social and strategic issues have to be planned prior to technological infrastructure. New rules of information processing and use still emerge and in interfaces information behavior and cognitive and affective ecological models will be mirrored. Real value of information repositories is in information and content. Information professionals and academic librarians are parts of information infrastructures. They will have to guard the meaningful information processes in the electronic information environment.

Acknowledgement

The paper was prepared as part of the research project KEGA 3/7275/09.

References

- BORGMAN, Christine. Scholarship in the Digital Age. *Information, Infrastructure and the internet*. London : MIT Press 2007. 336 s. ISBN 978-0-262-02619-2.
- CONNAWAY, Silipigni Lynn. Make room for the Millennials : Changes to systems and services will appeal to this generation. *NextSpace* [online]. October 2008, no.10 [cited 2009-06-06] p.18-19. Available from: <<http://www.oclc.org/nextspace/010/research.htm>>. ISSN 1559-0011.
- DAVENPORT, Thomas H., PRUSAK, Laurence. *Information Ecology: Mastering the Information and Knowledge Environment*. New York : Oxford Univ. Press, 1997. 255 s. ISBN 0-19-511168-0.
- DEMPSEY, Lorcan. Always on: Libraries in a World of Permanent Connectivity. *First Monday* [online]. Jan 2009, vol. 14, no. 1–5. [cited 2009-06-06]. Available from: <<http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/2291/2070>>.
- FAST, Karl V., SEDIG, Kamran. The INVENT framework: Examining the role of information visualization in the reconceptualization of digital libraries. *Journal of Digital Information* [online]. 2005, vol 6, no. 3 [cited 2009-02-02]. Available from: <<http://journals.tdl.org/jodi/article/view/66/69>>.
- GIVEN, Lisa M. Emotional Entanglements on the University Campus: The Role of Affect in Undergraduates' Information Behaviors. In: NAHL D. and BILAL D. (Eds.). *Information and Emotion: the Emergent Affective Paradigm in Information Behavior Research and Theory*. Medford, NJ: Information Today, 2007, p. 161-175.
- INFORMATION Behaviour of the Researcher of the Future*. A CIBER briefing paper [online]. University College of London. Centre for Information behaviour research, Jan 2008 [cited 2009-03-28] 35p. Available from: <http://www.jisc.ac.uk/media/documents/programmes/reppres/gg_final_keynote_11012008.pdf>.
- LYNCH, Clifford. The Institutional Challenges of Cyberinfrastructure and e-Research. *EDUCAUSE Review* [online]. Nov/Dec 2008, vol. 43, no.6 [cited 2009-03-16]. Available from: <<http://connect.educause.edu/Library/EDUCAUSE+Review/TheInstitutionalChallenge/47446>>.
- NAHL, Diane. A discourse analysis technique for charting the flow of micro-information behaviour. *Journal of Documentation*. 2007, vol. 63, no. 3, p. 323-339.
- NAHL, Diane and BILAL, Dania, (Eds.) *Information and Emotion: the Emergent Affective Paradigm in Information Behavior Research and Theory*. Medford, NJ: Information Today, 2007. 359 p. ISBN 978-1-57387-310-9.
- NAHL, Diane. The Centrality of the Affective in Information Behavior. In: NAHL D. and BILAL D. (Eds.). *Information and Emotion: the Emergent Affective Paradigm in Information Behavior Research and Theory*. Medford, NJ: Information Today, 2007, p. 3-37.
- NARDI, Bonnie A. and O'DAY, Vicki L. *Information Ecologies: Using Technology with Heart*. Cambridge, Mass.: MIT Press, 1999. 232 p. ISBN 0-262-14066-7.
- NICHOLAS, David, DOBROWOLSKI, Tom, WITHEY, Richard, RUSSELL, Chris, HUNTINGTON, Paul and WILLIAMS, Peter. Digital information consumers, players and purchasers: information seeking behaviour in the new digital interactive environment. *Aslib Proceedings*, 2003, vol. 55, no. 1/2, p. 23-31. ISSN: 0001-253X. DOI: 10.1108/00012530310462689.
- NICHOLAS, David, HUNTINGTON, Paul, WILLIAMS, Peter, DOBROWOLSKI, Tom. 2005. The Digital Information Consumer. Chapter 11. In: SPINK, A., Cole, C. (eds.). *New Directions in Human Information Behavior*. Dordrecht: Springer 2005, p. 203-228. ISBN 1-4020-3667-1.
- OBLINGER, Diana G. and OBLINGER, James L. (eds.). *Educating the Net Generation* [online]. Washington: EDUCAUSE, 2005 [cited 2009-06-06]. Available from: <<http://www.educause.edu/educatingthenetgen>>.
- NOVAK, Joseph D. and CAÑAS, Alberto J. *The Theory Underlying Concept Maps and How to Construct Them* [online]. Technical Report IHMC CmapTools 2006-01 Rev 01-2008, Florida Institute for Human and Machine Cognition, 2008 [cited 2009-06-06]. Available from: <<http://cmap.ihmc.us/Publications/ResearchPapers/TheoryUnderlyingConceptMaps.pdf>>.
- PRABHA, Chandra, CONNAWAY, Silipigni Lynn, OLSZEWSKI, Lawrence and JENKINS, Lillie R. What is enough? Satisficing information needs. *Journal of Documentation* [online] 2007, vol. 63, no. 1, p. 74-89. [Cited Feb 8, 2009] Pre-print available from: <<http://www.oclc.org/research/publications/archive/2007/prabha-satisficing.pdf>>.
- PRENSKY, Marc. Digital Natives, Digital Immigrants. *On the Horizon* [online], Oct 2001, vol. 9, no. 5, [cited 2009-06-06] p. 1-6. Article available from: <<http://www.marcprensky.com/writing/default.asp>>.
- RADFORD, Marie L., CONNAWAY, Silipigni Lynn. *Seeking Synchronicity: Evaluating Virtual Reference Services from User, Non-user, and Librarians perspectives* [online]. 2005. A research project funded by the National Leadership Grants for Libraries program of the Institute of Museum and Library Services (IMLS) Rutgers, The State University of New Jersey and OCLC, Inc. [cited 2009-02-08]. Available from: <<http://www.oclc.org/research/projects/synchronicity>>.
- SCHMITZ, Dawn. *The Seamless Cyberinfrastructure: The Challenges of Studying Users of Mass Digitization and Institutional Repositories*. [online]. Council on Library and Information Resources. April 2008 [cited 2009-06-06]. Available from: <www.clir.org/pubs/archives/schmitz.pdf>.
- SCHNEIDERMAN, Ben. Science 2.0. *Science* [online], 7 March 2008, vol. 319, no. 5868 [cited 2009-02-08] p. 1349-1350. Available from

- (restricted access): <<http://www.sciencemag.org/content/vol319/issue5868/index.dtl>>.
- SMITH, Abby. *New Model Scholarship: How Will it Survive?* [online]. Washington D.C.: Council on Library and Information Resources, March 2003 [cited 2009-05-15]. Available from: <<http://www.clir.org/pubs/reports/pub114/pub114.pdf>>. ISBN 1-887334-99-8.
- Sharing, Privacy and Trust in Our Networked World* [online]. OCLC membership report. Dublin, Ohio: OCLC Online Computer LibraryCenter, 2007 [cited 2009-05-15]. Available from: <<http://www.oclc.org/reports/privacyandtrust>> Also available in pdf version from: <<http://www.oclc.org/reports/pdfs/sharing.pdf>>. ISBN: 1-55653-370-5.
- STEINEROVÁ, Jela, GREŠKOVÁ, Mirka and ŠUŠOL, Jaroslav. Prieskum relevancie informácií: Výsledky analýz rozhovorov s doktorandmi FiFUK. [Survey of relevance. Research report.] Bratislava: CVTI, 2007. 150 p. ISBN 978-80-85165-93-7.
- STEINEROVÁ, Jela, ŠUŠOL, Jaroslav, MAKULOVÁ, Soňa, MATTHAEIDESOVÁ, Marta, VERČEKOVÁ, Jana and RANKOV, Pavel. *Správa o empirickom prieskume používateľov knižníc: ako súčasť grantovej úlohy VEGA 1/9236/02 Interakcia človeka s informačným prostredím v informačnej spoločnosti*. Bratislava: FiF UK KKIV, 2004. 113 p.
- STEINEROVÁ, Jela and ŠUŠOL, Jaroslav. Library users in human information behaviour. *Online Information Review* [online]. 2005, vol. 29, no. 2 [cited 2009-06-06] p. 139-156. Available from (restricted access): <<http://www.ingentaconnect.com/content/mcb/264/2005/00000029/00000002/art00002>>. ISSN 1468-4527.
- STEINEROVÁ, Jela and ŠUŠOL, Jaroslav. Users' information behaviour – a gender perspective. *Information Research* [online]. April 2007, vol. 12, no. 3, paper 320 [cited 2009-06-06]. Available from: <<http://informationr.net/ir/12-3/infres123.html>>. ISSN 1368-1613.
- STEINEROVÁ, Jela. *Informačné správanie: Pohľady informačnej vedy*. Bratislava: CVTI SR, 2005. 189 p. ISBN 80-85165-90-2.
- STEINEROVÁ, Jela. Relevance assessment for digital libraries. *Mousaion* [online]. 2007. vol. 25, no. 2 [cited 2009-06-06] p. 37-57. Available from (for authorized users): <http://journals.sabinet.co.za/ej/ejour_mousaion.html>. ISSN 0027-2639.
- STEINEROVÁ, Jela. Seeking relevance in the academic information use. *Information* [online]. November 2008, vol. 13, no. 4, paper 380. Proceedings of the 7th conference on Information Seeking in Context [ISIC], Vilnius, September 2008 [cited 2009-06-06]. Available from: <<http://informationr.net/ir/13-4/paper380.html>> ISSN 1368-1613. Also published: *Information seeking in context*. Vilnius: Vilnius university, 2008, p. 77-91.
- SOMPEL, Herbert Van de, PAYETTE, Sandy, ERICKSON, John and LAGOZE, Carl 2004. Rethinking Scholarly Communication. Building the System that Scholars Deserve. *D-Lib Magazine* [online]. Sep 2004, vol. 10, no. 9 [cited 2009-06-06]. Available from: <<http://www.dlib.org/dlib/september04/vandesompel/09vandesompel.html>>. ISSN 1082-9873.
- WILSON, T.D. A re-examination of information seeking behaviour in the context of activity theory. *Information Research* [online]. July 2006, vol. 11, no. 4, 2006, paper 260 [cited 2009-06-06]. Available from: <<http://InformationR.net/ir/11-4/paper260.html>>. ISSN 1368-1613.
- WILLIAMSON, Kirsty. Ecological Theory of Human Information Behavior. In FISHER, Karen E., ERDELEZ, Sanda and McKECHNIE, Lynne E.F. (eds.) *Theories of Information Behavior*. Medford: Information Today, 2005. p. 128-132. ISBN 1-57387-230-X.

Website resources

- DRIVER – Digital Repository Infrastructure Vision for European Research. EUROPEAN Commission [online]. [Cited 2009-03-17]. Available from: <<http://www.driver-repository.eu/>>.
- OXFORD e-Research Centre (OeRC) [online]. Oxford Internet Institute. [Cited 2009-03-17]. Available from: <<http://www.oerc.ox.ac.uk/>>.

Institutional Repositories and Document Citation

Linda Skolková & Miloslav Nič

`linda.skolkova@ff.cuni.cz`; `miloslav.nic@vscht.cz`

Institute of Information Studies and Librarianship, Faculty of Arts, Charles University, Prague, Czech Republic

Institute of Chemical Technology, Prague, Czech Republic

Abstract: The paper presents a brief overview of citation formats in relation to institutional repository software, i.e. its particular installations, and investigates a variety of possibilities how users can further work with citation data.

Four particular installations of repository software are included: DSpace@MIT (DSpace), E-LIS (EPrints), ARROW Repository (Fedora) and CERN Document Server (CDS Invenio). While DSpace@MIT and ARROW Repository provide their users with very limited citation export options or tools, E-LIS and CERN Document Server functionalities can serve as very good examples in this regard.

A brief survey of citation export capabilities of selected journals, digital libraries and citation databases has shown which formats are widely used – from among those we have chosen MODS, BibTeX and RIS when working with the repositories and citation managers. The particular experiments with citation data imports and exports have brought rather interesting results – at first sight, no major problems were encountered during imports and exports; however, when the imported and exported data were checked, we found out that some of the data were kept and some were due to transformations performed by one of the citation managers lost.

The whole paper is focused on open source software – both institutional repository software and citation managers we have worked with are open source. The use of open source technologies has not been chosen arbitrarily – we believe that the importance of open source – as a parallel to Open Access movement – in library and information services belongs to current trends which should not be underestimated. On the other hand, we still have to keep in mind that open source does not mean perfect, however, it makes it possible for basically anyone to contribute with a piece of code. This may serve as an inspiration for (active) users institutional repository software (or its installation) which currently do not offer sophisticated citation export capabilities (unlike a number of mostly commercial digital libraries, fee-based journals or citation databases). For those institutions or individuals who consider setting up a repository, citation export capabilities may become one of selection criteria.

Last but not least, when users are concerned, the need to develop effective ways of citing documents available in various versions (say in an eprint archive and in a journal) becomes apparent – for example when creating references according to ČSN ISO 690 and especially ČSN ISO 690-2, a field for notes may be used to capture the relationship between document versions.

Keywords: institutional repositories, document citation, citation formats, citation manager

Introduction

This paper focuses on comparison of citation export options offered by selected open source software tools used for institutional repositories. Furthermore, it presents the results of citation sample data import into selected citation managers. Citations of documents available from institutional repositories which are not explicitly included in ČSN ISO 690 and ČSN ISO 690-2 standards dealing with citations and references are also discussed.

The paper is organized in the following way: First, in the introductory part, terminology used in the paper is explained. It is followed by methodology, results, discussion and conclusions and references.

First, let us have a closer look at terminology which we shall operate with in this paper. We focus on citation formats – unlike *citation styles* (e.g. ČSN ISO 690, APA, MLA or Chicago citation styles) which prescribe how

bibliographic data are presented to human reader, *citation formats* (e.g. RIS, BibTeX and others mentioned in the paper) are primarily intended for machines - for automated processing performed by computers. Although citation formats typically contain detailed bibliographic data about a particular document (e.g. journal article, conference paper from proceedings etc.) which authors of new documents citing this document typically include as *references* at the end of their document and point to them from various parts of their document (thus creating *citations*), in this paper we prefer to use the adjective *citation* in relation to the formats. To make the text easy to read, we have also decided to use the term *citation manager* (instead of *bibliographic software*, *reference management software*, *reference manager* or similar terms) to denote software tools which authors use in order to process citations and references. More on related terminology can be found in Skolková L. Managing Citations and References with Software Tools.

Methodology

As to the methodology used, we proceeded in the following way:

First of all, we performed a survey of existing open source software tools which are suitable (or, to be more precise, used in practice) for institutional repositories. We consulted a variety of sources, e.g. *A Guide to Institutional Repository Software*,¹ *A Survey and Evaluation of Open-Source Electronic Publishing Systems*,² a list of software used by repositories included in Registry of Open Access Repositories (ROAR)³ or English Wikipedia entry on institutional repository⁴. We also took into account institutional repository software currently used in the Czech Republic.⁵ As a result of this survey, we chose four software tools.

Then – based on the lists of projects available at the software’s websites – we chose one particular installation per software to use in further steps of our research. Where necessary, we also studied the documentation, manuals and other available materials (such as tutorials) related to each software (available on the software’s official website). We focused on the software’s capabilities to export citation data so that these data can then be automatically processed by other information systems, especially by citation managers. In other words, we did not focus on export and import features from the system administrator’s (or repository content manager’s) point of view – instead, end user experience was our focal point.

After the survey of selected repository installations we assessed selected journals, digital libraries and citation databases which are widely used by scientists. The reason for this step is simple – these (mostly commercial or fee-based) resources are well known for the provision of various value-added services and functionalities for their users – including citation data processing. The results were very helpful in terms of the next steps in our research – choice of citation managers and citation formats to work with.

As indicated, our choice of open source citation managers was based on demands which scientists – authors are likely to come up with – when working with digital information resources, especially (individual) journals, digital libraries and citation databases. Apart from this input, we also used a list of available citation managers in

¹ *A Guide to Institutional Repository Software* <http://www.soros.org/openaccess/pdf/OSI_Guide_to_IR_Software_v3.pdf>.

² CYZYK M. and CHOUDHURY S. *A Survey and Evaluation of Open-Source Electronic Publishing Systems*

³ *Registry of Open Access Repositories* <<http://roar.eprints.org/>>.

⁴ *Institutional Repository* <http://en.wikipedia.org/wiki/Institutional_repository>.

⁵ As presented in
KREJČÍŘ V. *Systémy pro tvorbu digitálních knihoven*
RYGLOVÁ P. *Zpráva z 2. setkání českých uživatelů systému DSpace*
LHOTÁK M. *Open source pro digitální knihovnu*
or *Národní úložiště šedé literatury*

Wikipedia.⁶ From a number of options, we eventually chose two citation managers and three widely used formats to proceed with.

Then we exported a sample set of citations from two repositories (we chose two of the four particular installations surveyed – those which were well capable of exporting citation data) and imported it into the citation manager. Afterwards, we assessed the import and exported the data again so that it can be imported to the second citation manager and compared with the other available data sets.

Last but not least, based on citation data we have worked with both in this research and in other projects, we briefly introduce the problem of citing documents from institutional repositories which are in a certain relation to other published documents.

Results

Based on the resources cited in the previous part of this paper, we have chosen four institutional repository software tools – DSpace, EPrints, Fedora and CDS Invenio.⁷ All of these fall into the realm of open source software and belong to widely used software. As of May 11, 2009, at least 565 repositories used DSpace⁸, at least 269 archives were powered by EPrints⁹, at least 146 projects used Fedora¹⁰ and at least 24 projects used CDS Invenio¹¹. From among the projects/repositories, four have been chosen – DSpace@MIT (powered by DSpace, as the title indicates), E-LIS (powered by EPrints), ARROW Repository (powered by Fedora) and CERN Document Server (powered by CDS Invenio).

As we will show in the next sections, while DSpace@MIT and ARROW Repository provide their users with very limited citation export options or tools, E-LIS and CERN Document Server functionalities can serve as very good examples in this regard.

⁶ *Comparison of reference management software* <http://en.wikipedia.org/wiki/Comparison_of_reference_management_software>.

⁷ *DSpace* <<http://www.dspace.org/>>.

Open Access and Institutional Repositories with EPrints <<http://www.eprints.org/>>.

Fedora Commons <<http://fedora-commons.org/>>.

CDS Invenio Overview <<http://cdsware.cern.ch/invenio/index.html>>.

⁸ *DSpace, Repository list* <<http://www.dspace.org/index.php/DSpace-Instances/Repository-List.html>>.

⁹ *Open Access and Institutional Repositories with EPrints. Sites powered by EPrints* <<http://www.eprints.org/software/archives/>>.

¹⁰ *Fedora Commons. Fedora Commons Community Registry* <<https://fedora-commons.org/confluence/display/FEDINFO/Fedora+Commons+Community+Registry>>.

¹¹ *CDS Invenio Demo* <<http://cdsware.cern.ch/invenio/demo.html>>.



Figure 1 — Simple record from DSpace@MIT with citable URI.¹²

DSpace: DSpace@MIT

In case of DSpace@MIT¹³ no particular citation data export formats are offered. When the user performs a search, he or she can view the individual search result (record representing an item – or a digital object – in the repository and eventually the item itself). Two options are offered – simple (see Figure 1) and full record. Simple record provides bibliographic data related to the item with labels in natural language (e.g., Title, Author or Publisher) while the full record provides more detailed information with Dublin Core element labels (such as dc.title, dc.contributor.author or dc.publisher). Both record types provide a URI (Uniform Resource Identifier), in this case represented by handle, which serves as a persistent identifier of the item. However, it is not possible to export citation data in a format which can be processed automatically by citation managers. The citation export option is offered neither for individual records, nor for record sets (if the user tried to work with more than search result at a time).

EPrints: E-LIS

E-LIS eprint archive¹⁴ uses EPrints and enables users to export citation data in a variety of formats. When search

results are presented to the user (see Figure 2), he or she can export the whole set (but not a particular subset) to thirteen different formats: ASCII Citation, BibTeX, Dublin Core, EP3XML, EndNote, Eprints Application Profile, HTML Citation, ISO Citation, METS, Object IDs, OpenURL Context Object, Refer and Reference Manager. When the user works with an individual (simple) record, then sixteen metadata export options are included – apart from those already mentioned above, Full Metadata, DIDL and Simple Metadata are offered. When full record (all metadata) are viewed, then citation options are not directly available (metadata is presented in a table).



Figure 2 — Search results display from E-LIS eprint archive with export options applicable to the the whole set.¹⁵

Fedora: ARROW Repository

When the user performs a search in the ARROW Project repository at Monash University¹⁶, no options to export citations are offered, so it is not possible to export either citations representing a set of records (or, more precisely, search results), or – when the individual record is displayed (see Figure 3) – the citation representing the record in question. The user is only able to copy the identifier (handle), a single citation in a specific citation style or identifier of the item within the repository. Basic (though not complete) bibliographic data are also shown in the upper part of the page (above the record itself).

¹² Source: <<http://dspace.mit.edu/handle/1721.1/30592>> [captured 2009-05-12].

¹³ Welcome to DSpace at MIT <<http://dspace.mit.edu/>>.

¹⁴ E-LIS <<http://eprints.rclis.org/>>.

¹⁵ Source: <http://eprints.rclis.org/cgi/search/simple?q=digital+libraries+software&q2=&order=byyear&q_merge=ALL&q2_merg=ALL&_action_search=Submit> [captured 2009-05-12].

¹⁶ Monash University Arrow Repository <<http://arrow.monash.edu.au/vital/access/manager/Index>>.

MONASH University
ARROW Repository

Home | Browse | Contact | Help |

digital libraries software

Search

Advanced Search Show All 4036

Home > List Of Titles > The ARROW project: a consortial institutional repository solution, combining open source and proprietary software

Groenewegen, David, Treloar, Andrew, . Emerald Group Publishing Limited; 2008. The ARROW project: a consortial institutional repository solution, combining open source and proprietary software.

Please use this identifier to cite or link to this item:
<http://arrow.monash.edu.au/hdl/1959.1/43965>

Title **The ARROW project: a consortial institutional repository solution, combining open source and proprietary software**

Creator Groenewegen, David

Creator Treloar, Andrew

Date 2008

Figure 3 — A record from ARROW repository with identifier to be used when citing the item.¹⁷

CDS Invenio: CERN Document Server

When CERN Document Server¹⁸ presents search results, users are able to choose from eleven output formats: Excel, HTML brief, BibTeX, HTML address label, HTML detailed, HTML MARC, HTML photo captions only, HTML portfolio, MODS, XML Dublin Core, XML MARC (see Figure 4 and Figure 5). It is also possible to work with record subsets – each search result can be ticked in a check-box.

Registered users are allowed to use so called Baskets – the user can store either a whole set or a selected subset of search results in the basket. When a particular record is displayed (see Figure 6) – either directly from the search results page, or from the basket – the user is able to export citation data in six formats. These are BibTeX, MARC, MARCXML, DC, EndNote and NLM.

CERN Document Server

CDS | Indico | Library | Bulletin | EDM3

Hledej | Přidej | Návoděda | Your CDS

Hlavní stránka > Výsledky hledání

CERN Document Server

Hledej:

digital libraries software

všechna pole

Hledej Prolistuj

Tipy pro vyhledávání · Rozšířené Hledání

Hledej kolekce:

*** všechny kolekce ***

Sort: Display results: Output format:

poslední záznam nejdříve

10 výsledky

HTML brief

sest: seskupené podle kolekcí

- nebo seřadit podle -

Results overview: Found 56 records in 0.02 seconds.

Articles & Preprints 47 records found

Books & Proceedings 2 records found

Periodicals & Progress Reports 1 records found

Multimedia & Outreach 7 records found

Articles & Preprints 47 records found 1 - 10 >>> jump to record: 1

Figure 4 — Search results display from CERN Document Server with output format options – in this case HTML brief is applied.¹⁹

```
<!-- Search-Engine-Total-Number-Of-Results: 56 -->
- <collection>
- <record>
  <controlfield tag="001">1174561</controlfield>
  <controlfield tag="003">SzGeCERN</controlfield>
  <controlfield tag="005">20090505105107.0</controlfield>
  - <datafield tag="037" ind1=" " ind2=" " >
    <subfield code="a">CERN-THESIS-2009-032</subfield>
  </datafield>
  - <datafield tag="041" ind1=" " ind2=" " >
    <subfield code="a">eng</subfield>
  </datafield>
  - <datafield tag="100" ind1=" " ind2=" " >
    <subfield code="a">Raae, L C</subfield>
    <subfield code="u">Bergen University College</subfield>
  </datafield>
  - <datafield tag="245" ind1=" " ind2=" " >
```

Figure 5 — Search results display from CERN Document Server in MARCXML format.²⁰

¹⁷ Source: <<http://arrow.monash.edu.au/hdl/1959.1/43965>> [captured 2009-05-12].

¹⁸ CERN Document Server <<http://cdsweb.cern.ch/>>.

¹⁹ Source: <http://cdsweb.cern.ch/search?ln=cs&p=digital+libraries+software&f=&action_search=Hledej&c=CERN+Document+Server&sf=&so=d&rm=&rg=10&sc=1&of=hb> [captured 2009-05-12].

²⁰ Source: <http://cdsweb.cern.ch/search?ln=cs&p=digital+libraries+software&f=&action_search=Hledej&c=CERN+Document+Server&sf=&so=d&rm=&rg=10&sc=1&of=xm> [captured 2009-05-12].

Information | References | Discussion | Usage statistics | Plný text

Thesis

Report number: CERN-THESIS-2009-032

Title: **Interactive Editing and Cataloging Interfaces for Modern Digital Library Systems**

Author(s): [Raae, L C](#) (Bergen University College) ; [Helstrup, H](#) (dir.) (Bergen University College)

Imprint: Bergen : Bergen University College, 2009. - 62 p.

Thesis note: Thesis : Bergen, Bergen University College : 2009

Note: Presented on 14 Apr 2009

Abstract: The next-generation High Energy Physics information system, INSPIRE, is being built by combining the content from the successful SPIRES database of bibliographic information with the CDS Invenio software being developed at CERN, an open-source platform for large digital library systems. The project is a collaboration between four major particle physics laboratories in Europe and the U.S. New tools are being developed to enable the global cooperation between catalogers at these labs. The BibEdit module will provide a central interface for the editing, enrichment, correction and verification of a record on its way into the system, by processing and presenting data from several supporting modules to the cataloger. The objective is to minimize the time and actions needed by the cataloger to process the record. To create a fast and powerful web application we make use of modern AJAX technology to create a dynamic and responsive user interface, where server communication happens in the background without delaying the work of the user. On the Apache web server a Python engine processes the different types of requests and submits persistent changes to a MySQL database, while caching temporary data on file. Data is transferred between the client and server using the JSON protocol. By using JavaScript, jQuery and jQuery plug-ins we enable such clientside UI features as in-place editing, hotkeys, autosuggest etc.

Email contact: lars.christian.raae@cern.ch

Record created 2009-04-29, last modified 2009-05-06 [Podobné záznamy](#)

Fulltext: Rate this document: Add to personal basket
Export as [BibTeX](#), [MARC](#), [MARCXML](#), [DC](#), [EndNote](#), [NLM](#)

(Not yet reviewed)

Figure 6 — A full record display from CERN Document Server in with six export formats.²¹

Citation formats in scientific community

In the previous parts of the paper we have dealt with citation export options at selected repositories. Now let us have a look at some of the leading journals and digital libraries and also at citation databases so that we can have a broader overview of formats which are commonly used in scientific community.

Nature²², a high impact-factor journal (28.751), offers users to save citations into Connotea and also to export them in RIS format.

Science²³, another journal with high impact factor (26.372), provides several formats in which the user can download citations. These are EndNote, Reference Manager

(RIS), ProCite (RIS), BibTeX, RefWorks and Medlars. It is also possible to post the item to CiteULike or to view PubMedCitation. Last but not least the users can request an additional format be added.

SpringerLink²⁴, a digital library operated by Springer, provides users with two export options – the article (or, to be precise, its bibliographic data) can be exported in RIS or in text format.

ScienceDirect²⁵, Elsevier's digital library, also offers RIS format (it informs the users that this format can be imported into Reference Manager, ProCite and EndNote), RefWorks Direct Export, ASCII format and BibTeX format. Furthermore, it is possible to add items into 2collab system.

Wiley InterScience²⁶, Wiley (and Blackwell Synergy) digital library, offers citation download in two formats – Plain Text and EndNote.

Scopus²⁷ as an abstract and citation database offers RIS format (Reference Manager, ProCite, EndNote), Text (ASCII format), RefWorks Direct Export and Comma Separated File, .csv (e.g. Excel). It also offers BibTeX – however, this format is not included in the format list but can be found among citation styles. The output options are the same both in the case of search results display and in the case of individual records. Individual records also provide an extra export option – users can use 2collab bookmarks.

Web of Science, now part of a wider platform called ISI Web of Knowledge²⁸, lets users export citation data directly to EndNote Web. Other options include export to EndNote/RefMan/ProCite (in this case RIS format is not used; instead, Web of Science uses its own format – data are saved in CGI format, see *Figure 7*), to HTML, Plain Text and also to tab-delimited files (for Win and for Mac).

```

FN ISI Export Format
VR 1.0
PT J
AU Stewart, M
AF Stewart, Mandy
TI Resource Discovery at the British Library
SO INTERLENDING & DOCUMENT SUPPLY
SN 0264-1615
PY 2009
VL 37
IS 1
BP 46
EP 48
DI 10.1108/02641610910938622
UT ISI:000264262700008
ER
EF

```

Figure 7 — ISI Export Format.²⁹

²¹ Source: <<http://cdsweb.cern.ch/record/1174561?ln=cs>> [captured 2009-05-12].

²² Nature <<http://www.nature.com/nature/index.html>> .

²³ Science <<http://www.sciencemag.org/>>.

²⁴ SpringerLink Home <<http://springerlink.com/home/main.mpx>>.

²⁵ ScienceDirect <<http://www.sciencedirect.com/>>.

²⁶ Wiley InterScience <<http://www3.interscience.wiley.com/cgi-bin/home>>.

²⁷ Scopus <<http://www.scopus.com/>>.

²⁸ ISI Web of Knowledge <<http://isiknowledge.com/>>.

²⁹ Source: Web of Science export file in CGI format viewed in PSPad editor [captured 2009-05-12]

Formats	URLs
BibTeX	http://www.bibtex.org/
Dublin Core	http://dublincore.org/
MARCXML	http://www.loc.gov/marc/marcxml.html
METS	http://www.loc.gov/standards/mets/
MODS	http://www.loc.gov/standards/mods/
RefWorks (tagged format)	http://refworks.scholarsportal.info/Refworks/help/RefWorks_Tagged_Format.htm
RIS	http://www.refman.com/support/risformat_intro.asp

Table 1 — Overview of selected citation format specifications.

For readers more interested in citation format specifications, we have compiled an overview of selected widely used citation formats (see *Table 1*; data in the table are ordered alphabetically).

Based on this brief overview of citation formats and their use in selected resources (journals, digital libraries and citation databases), we have decided to focus on two widely used formats – MODS and BibTeX (see *Figure 8* and *9* for particular examples of citation data in these formats) when imports from EPrints (E-LIS) and CDS Invenio (CERN Document Server) are considered and RIS (see *Figure 10*) and BibTeX when it comes to exports from citation managers.

Therefore in the next section of the paper we present two open source citation managers – Zotero and Connotea³⁰ – and the results of sample citation data imports into these citation managers. They have been chosen for two basic reasons – 1) they belong to the realm of open source software and the user is not charged any fees for using them in practice, and 2) both are capable of working with a variety of import and export formats.

```

<mods:mods version="3.0"
  xsi:schemaLocation="http://www.loc.gov/mods/v3
  http://www.loc.gov/standards/mods/v3/mods-3-0.xsd">
  <mods:titleInfo>
    <mods:title>
      Institutional Repositories, Open Source Options, and Libraries
    </mods:title>
  </mods:titleInfo>
  <mods:name type="personal">
    <mods:namePart type="given">Megan E.</mods:namePart>
    <mods:namePart type="family">Amaral</mods:namePart>
  </mods:name>
  <mods:role>
    <mods:roleTerm type="text">author</mods:roleTerm>
  </mods:role>
  <mods:abstract>
    This paper examines the purposes of institutional digital
    repositories and the development options that are available via
    open source software. Preservation of an institution's digital
    output is valuable to the submitting authors and the institution.
    There are digital repository products available as open source
    software that may be extremely helpful to an organization
    establishing an institutional repository. The role of libraries
    and librarians in this process is discussed.
  </mods:abstract>
  <mods:classification authority="lcc">HS. Repositories.

```

Figure 8 — Part of a sample citation in MODS format from E-LIS as viewed in Mozilla Firefox browser.³¹

³⁰ Zotero <<http://www.zotero.org/>>.

Connotea <<http://www.connotea.org/>>.

³¹ Source: <<http://eprints.rclis.org/cgi/export/15355/MODS/eprints3-eprint-15355.xml>> [captured 2009-05-12].

```

@misc{eprints315355,
  volume = {1},
  number = {1},
  month = {December},
  author = {Megan E. Amaral},
  title = {Institutional
  Repositories, Open Source Options, and
  Libraries},
  publisher = {Ellyssa Kroski},
  journal = {Open and Libraries Class
  Journal},
  year = {2008},
  keywords = {Digital repositories,
  Institutional repositories, open source
  software, DSpace},
  url = {http://eprints.rclis.org/
  15355/},
  abstract = {This paper examines the
  purposes of institutional digital
  repositories and the development options that
  are available via open source software.
  Preservation of an institution's digital
  output is valuable to the submitting authors
  and the institution. There are digital
  repository products available as open source
  software that may be extremely helpful to an
  organization establishing an institutional
  repository. The role of libraries and
  librarians in this process is discussed.}
}

```

Figure 9 — Sample citation in BibTeX format from E-LIS as viewed in PSPad editor.³²

```

TY - CHAPTER
JF - Preserving Digital Information
T1 - Everyday Digital Content
Management
SP - 181
EP - 204
PY - 2007///
UR - http://dx.doi.org/10.1007/978-3-
540-37887-7_9
M3 - 10.1007/978-3-540-37887-7_9
N2 - For the first thirty years of
digital preservation, archives managed
their digital collections with ... a
simple storage system and ... a
catalog database. Although the
fundamental design of a digital
archive system has remained the same .
.. a contemporary digital archive
needs more than [storage] for magnetic
tapes and a [database] for the
catalog. The rapid growth of digital
material in both volume and complexity
[and] rising expectations of archives
users ... have all contributed to the
redefinition of digital archive
functions.
ER -

```

Figure 10 — Sample citation in RIS format from SpringerLink as viewed in PSPad editor.³³

³² Source: <<http://eprints.rclis.org/cgi/export/15355/BibTeX/eprints3-eprint-15355.bib>> [captured 2009-05-12].

³³ Source: doi: 10.1007/978-3-540-37887-7 [captured 2009-05-12]

Zotero

Zotero offers seven import formats (used when importing data from a local file; see *Figure 11*) and seven export formats (see *Figure 12*). When the two sets are compared, four formats used in both cases can be identified – these are RIS, MODS, Refer/BibIX and BibTeX. Zotero – as a web browser plugin – also enables other ways of inserting citation data. It is possible to insert the data manually or capture the data representing a web page automatically and save them into this citation manager. To achieve this, all users need to do is to click on a small icon in the browser's address bar.

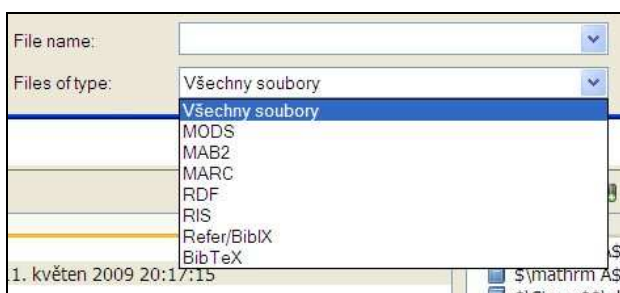


Figure 11 — Zotero import formats.³⁴

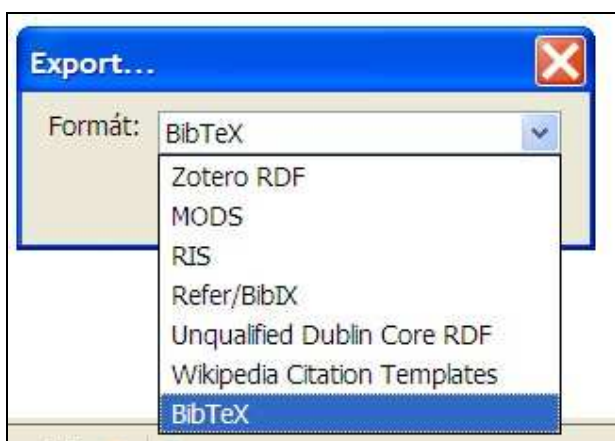


Figure 12 — Zotero export formats.³⁵

Connotea

Connotea is capable of working with seven import formats (intended for uploading data from a locally stored file) – these are RIS, BibTeX, EndNote, MODS, ISI Web of Knowledge and Firefox bookmarks (see *Figure 13*). As to export formats (see *Figure 14*), six are highlighted as the most important ones (RIS, EndNote, BibTeX, MODS (XML), Word 2007 Bibliography and Simple Text Citations), while three other formats (RSS, RDF and Plain) enable user to process data yet in a different way. Users can also insert data by clicking on Connotea browser button, by using a form or DOIs (Digital Object Identifiers).

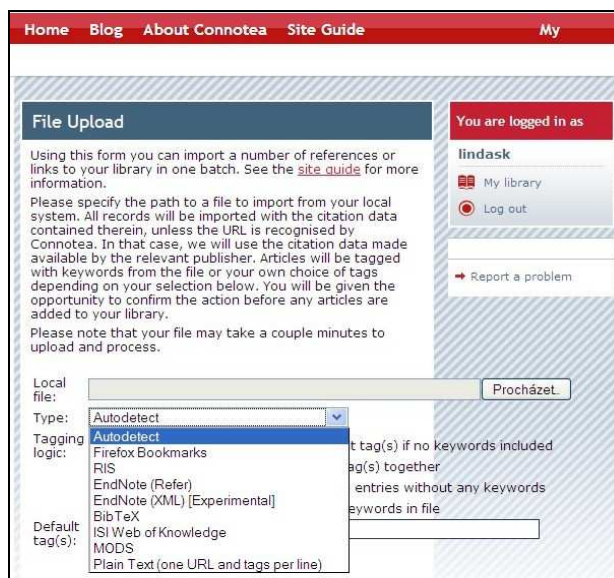


Figure 13 — Connotea import formats.³⁶

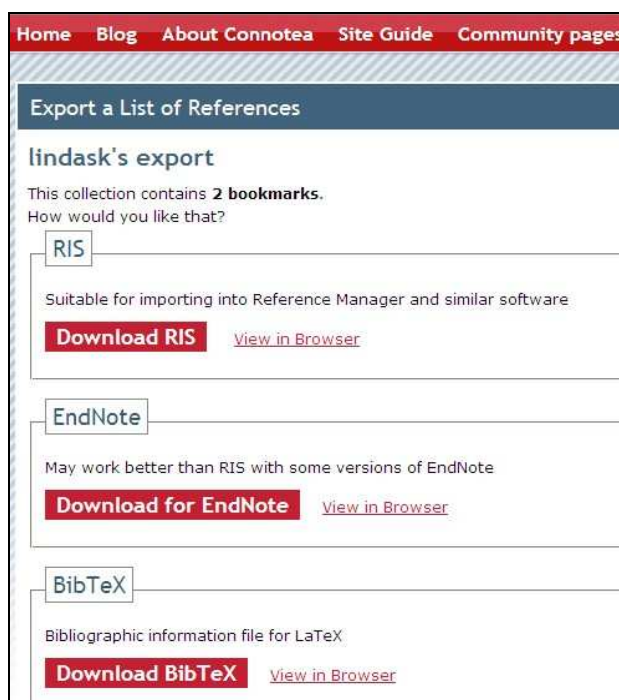


Figure 14 — Some of Connotea export formats.³⁷

Citation data imports

We gathered a sample set of citations from E-LIS eprint archive and from CERN Document Server.

The E-LIS set comprised of 110 records representing documents on the topic of institutional repositories – these were exported into BibTeX format and then imported to Zotero.

³⁴ Source: Zotero plugin in Firefox [captured 2009-05-12].

³⁵ Source: Zotero plugin in Firefox [captured 2009-05-12].

³⁶ Source: <<http://www.connotea.org/upload/>> [captured 2009-05-12].

³⁷ Source: <<http://www.connotea.org/export/>> [captured 2009-05-12].

The CERN Document Server set comprised of 80 records which represented documents on digital libraries. The records were exported into MODS format and then also imported imported to Zotero.

At first sight, these initial steps proceeded without any significant problems. However, minor problems occurred when data was viewed in Zotero, e.g. a record representing a document written in Greek which contained characters from Greek alphabet could not be seen in Greek alphabet (see *Figure 15 to 17*).

Ανάπτυξη ψηφιακής βιβλιοθήκης και ιδρυματικού αποθετηρίου στην Κεντρική Βιβλιοθήκη ΕΜΠ

Κουλουρίης, Α. and Κόκκινος, Δ. and Αναγνωστόπουλος, Α. *Ανάπτυξη ψηφιακής βιβλιοθήκης και ιδρυματικού αποθετηρίου στην Κεντρική Βιβλιοθήκη ΕΜΠ.*, 2007
 - In 16ο Πανελλήνιο Συνέδριο Ακαδημαϊκών Βιβλιοθηκών, Πειραιάς (GR), 1-3 Οκτωβρίου 2007. [Conference Paper]

Full text available as:

 PDF - Requires a PDF viewer such as GSview, Xpdf or Adobe Acrobat Reader
 385Kb Language: Greek

Alternative Locations: http://dlib.ionio.gr/pubs/16o_koulouris_kokkinos.pdf,
<http://dspace.lib.ntua.gr/handle/123456789/546>

Figure 15 — A record representing a paper in Greek.³⁸

```
@misc{eprints311609,
  editor = {$\mathrm{A}\nu\sigma\tau\omicron\upsilon\pi\omicron\upsilon\lambda\omicron\varsigma$,
  title = {$\mathrm{A}\nu\alpha\pi\tau\upsilon\chi\eta$ \psi\eta\phi\iota\alpha\kappa\eta\varsigma$ \beta\iota\beta\lambda\iota\omicron\theta\eta\kappa\eta\varsigma$ \k\alpha\iota$ \iota\delta\upsilon\mu\alpha\tau\iota\kappa\omicron\upsilon$ \alpha\pi\omicron\theta\eta\tau\iota\rho\iota\omicron\upsilon$ \sigma\tau\eta\mu$ \k\epsilon\upsilon\tau\upsilon\kappa\eta$ \beta\iota\beta\lambda\iota\omicron\theta\eta\kappa\eta$ \epsilon\mu\pi$},
  author = {$\mathrm{A}\nu\alpha\gamma\omega\sigma\tau\omicron\upsilon\pi\omicron\upsilon\lambda\omicron\varsigma$, \mathrm{A}\nu\alpha\pi\tau\upsilon\chi\eta$ \psi\eta\phi\iota\alpha\kappa\eta\varsigma$ \beta\iota\beta\lambda\iota\omicron\theta\eta\kappa\eta\varsigma$ \k\alpha\iota$ \iota\delta\upsilon\mu\alpha\tau\iota\kappa\omicron\upsilon$ \alpha\pi\omicron\theta\eta\tau\iota\ρ\iota\omicron\upsilon$ \sigma\tau\eta\mu$ \k\epsilon\upsilon\tau\upsilon\kappa\eta$ \beta\iota\β\lambda\iota\omicron\theta\eta\kappa\eta$ \epsilon\mu\pi$},
  year = {2007},
  pages = {378--392},
  keywords = {????? ???? ???? ???? ?}}
```

Figure 16 — Part of the record representing a paper in Greek in BibTeX as viewed in PSPad editor.³⁹

³⁸ Source: <<http://eprints.rclis.org/11609/>> [captured 2009-05-13].

³⁹ Source: BibTeX export from E-LIS [captured 2009-05-13].

Νάzev: $\mathrm{A}\nu\sigma\tau\omicron\upsilon\pi\omicron\upsilon\lambda\omicron\varsigma$, $\mathrm{A}\nu\alpha\pi\tau\upsilon\chi\eta$ $\psi\eta\phi\iota\alpha\kappa\eta\varsigma$ $\beta\iota\beta\lambda\iota\omicron\theta\eta\kappa\eta\varsigma$ $\kappa\alpha\iota$ $\iota\delta\upsilon\mu\alpha\tau\iota\kappa\omicron\upsilon$ $\alpha\pi\omicron\theta\eta\tau\iota\ρ\iota\omicron\upsilon$ $\sigma\tau\eta\mu$ $\kappa\epsilon\upsilon\tau\upsilon\kappa\eta$ $\beta\iota\beta\lambda\iota\omicron\theta\eta\kappa\eta$ $\epsilon\mu\pi$

- ▼ **Editor:** $\mathrm{A}\nu\alpha\gamma\omega\sigma\tau\omicron\upsilon\pi\omicron\upsilon\lambda\omicron\varsigma$, $\mathrm{A}\nu\alpha\pi\tau\upsilon\chi\eta$ $\psi\eta\phi\iota\alpha\kappa\eta\varsigma$ $\beta\iota\β\lambda\iota\omicron\theta\eta\kappa\eta\varsigma$ $\kappa\alpha\iota$ $\iota\delta\upsilon\mu\alpha\tau\iota\kappa\omicron\upsilon$ $\alpha\pi\omicron\theta\eta\tau\iota\ρ\iota\omicron\upsilon$ $\sigma\tau\eta\mu$ $\kappa\epsilon\upsilon\tau\upsilon\kappa\eta$ $\beta\iota\β\lambda\iota\omicron\theta\eta\kappa\eta$ $\epsilon\mu\pi$
- ▼ **Author:** $\mathrm{A}\nu\alpha\gamma\omega\sigma\tau\omicron\upsilon\pi\omicron\upsilon\lambda\omicron\varsigma$, $\mathrm{A}\nu\alpha\pi\tau\upsilon\chi\eta$ $\psi\eta\phi\iota\alpha\kappa\eta\varsigma$ $\beta\iota\β\lambda\iota\omicron\theta\eta\kappa\eta\varsigma$ $\kappa\alpha\iota$ $\iota\delta\upsilon\mu\alpha\tau\iota\kappa\omicron\upsilon$ $\alpha\pi\omicron\theta\eta\tau\iota\ρ\iota\omicron\upsilon$ $\sigma\tau\eta\mu$ $\kappa\epsilon\upsilon\tau\upsilon\kappa\eta$ $\beta\iota\β\lambda\iota\omicron\theta\eta\kappa\eta$ $\epsilon\mu\pi$
- ▼ **Author:** $\mathrm{A}\nu\alpha\gamma\omega\sigma\tau\omicron\upsilon\pi\omicron\upsilon\lambda\omicron\varsigma$, $\mathrm{A}\nu\alpha\pi\tau\upsilon\chi\eta$ $\psi\eta\phi\iota\alpha\kappa\eta\varsigma$ $\beta\iota\β\lambda\iota\omicron\theta\eta\kappa\eta\varsigma$ $\kappa\alpha\iota$ $\iota\delta\upsilon\mu\alpha\tau\iota\kappa\omicron\upsilon$ $\alpha\pi\omicron\theta\eta\tau\iota\ρ\iota\omicron\upsilon$ $\sigma\tau\eta\mu$ $\kappa\epsilon\upsilon\tau\upsilon\kappa\eta$ $\beta\iota\β\lambda\iota\omicron\theta\eta\kappa\eta$ $\epsilon\mu\pi$
- ▼ **Author:** $\mathrm{A}\nu\alpha\gamma\omega\sigma\tau\omicron\upsilon\pi\omicron\upsilon\lambda\omicron\varsigma$, $\mathrm{A}\nu\alpha\pi\tau\upsilon\chi\eta$ $\psi\eta\phi\iota\alpha\kappa\eta\varsigma$ $\beta\iota\β\lambda\iota\omicron\theta\eta\kappa\eta\varsigma$ $\kappa\alpha\iota$ $\iota\delta\upsilon\mu\alpha\tau\iota\kappa\omicron\upsilon$ $\alpha\pi\omicron\theta\eta\tau\iota\ρ\iota\omicron\upsilon$ $\sigma\tau\eta\mu$ $\kappa\epsilon\upsilon\tau\upsilon\kappa\eta$ $\beta\iota\β\lambda\iota\omicron\theta\eta\kappa\eta$ $\epsilon\mu\pi$

(...) **Abstrakt:** This paper describes the Digital Library (DL) and Institution

Figure 17 — Part of the record representing a paper in Greek as viewed in Zotero.⁴⁰

Both data sets were then exported from Zotero in RIS format and imported in this format into Connotea. Generally speaking, this resulted in a successful import. However, there were also certain problems associated with the import. At this point it is worth noting that Connotea actually provides the user with explicit comments on the problems and indicates how they have been handled (see *Figure 18*).

Adding record #1 Include

The keyword "Repositorios tem?ticos" will become tag "Repositorios temticos" to accommodate tag naming rules. The link associated with this record has been understood and authoritative bibliographic information will be used in place of the data in the RIS file.

[Viabilidad de repositorios de biomedicina y ciencias de la salud en la Comunidad Valenciana](#)
 Maria-Franisca Abad-Garcia et al.
 To be posted by **lindask** to "**acceso abierto**" "**Biomedicina y ciencias de la salud**" "**Biomedicine and health sciences**" "**Comunidad Valenciana**." "**Institutional repositories**" "**open access**" "**Repositorios Institucionales**" "**Repositorios temticos**" **Spain** "**Subject repositories**" "**Thematic repositories**" "**Valencian Community**"

Adding record #2 Include

The link associated with this record has been understood and authoritative bibliographic information will be used in place of the data in the RIS file.

[Verzahnung von Jahresbibliographie und Institutional Repository der Universit?at des Saarlandes](#)
 Verzahnung von Jahresbibliographie und Institutional Repository der Universit?at des Saarlandes
 Gerhard Abel et al.
 To be posted by **lindask** to **bibliography** "**digital library**" "**Institutional Repository**" "**open access**"

Figure 18 — Changes in data performed by Connotea.⁴¹

⁴⁰ Source: BibTeX import into Zotero from E-LIS [captured 2009-05-13].

⁴¹ Source: <<http://www.connotea.org/upload>> [captured 2009-05-12].

The data set was then exported again, this time from Connotea into BibTeX format. We then compared sample records the original BibTeX data with data gathered again after two conversions. We can conclude that the data sets – from the data point of view – were rather different – e.g. the Connotea export file did not contain abstract, also other data (volume, number, pages or journal in which the contribution was originally published) were lost. Author and title data as well as year, keywords and URL were still present. This clearly illustrates transformations which took place while uploading data from Zotero to Connotea and also a different focus of Connotea (sharing and tagging data – in Connotea, keywords play a far more important role than in other citation tools).

Citing documents in institutional repositories from information point of view

The above mentioned problem – data loss – encountered when performing data exports and imports in various formats actually gives us an idea about the importance of completeness of citation (or bibliographic) data representing a particular document.

If we consider one particular citation standard – ČSN ISO 690 and especially ČSN ISO 690-2 (or ISO 690 and ISO 690-2, standards in original versions), we can say that citation issues related to the situation when there is say a published journal article whose version (preprint, postprint, published version or some other derived version) is deposited in a repository (either a repository at a particular institution or in a subject-oriented repository without institutional limits) and the article is also published in a respected journal are not covered (also due to the time when the standards were published). However, there is a field for notes regarding the cited document, so this field can be filled with necessary data. This way it could be ensured that the reader of the document in which the “problematic” document is cited is able to distinguish the real document studied by the author.

At this point, a question may arise as to why is all this so important. We would argue that nowadays when information and communication technologies enable easy information sharing, when Open Access movement has achieved significant attention and especially practical results, when it is often easy to locate fulltext via general or specialized search engines, there is a growing problem how to recognize whether a certain document found “somewhere on the Web” is actually an exact copy of say a journal article. As this task is not easy to resolve, we believe that we should avoid citing “authoritative” versions of documents (e.g. journal articles) while actually using their “unauthoritative” versions. On the other hand, it is useful to indicate the original source of the document we are about to cite.

Discussion and conclusions

We have presented a brief overview of citation formats in relation to institutional repository software, i.e. its particular installations, and a variety of possibilities how the users can further work with citation data.

Four particular installations of repository software are included: DSpace@MIT (DSpace), E-LIS (EPrints), ARROW Repository (Fedora) and CERN Document Server (CDS Invenio). While DSpace@MIT and ARROW Repository provide their users with very limited citation export options or tools, E-LIS and CERN Document Server functionalities can serve as very good examples in this regard.

A brief survey of citation export capabilities of selected journals, digital libraries and citation databases has shown which formats are widely used – from among those we have chosen MODS, BibTeX and RIS when working with the repositories and citation managers. The particular experiments with citation data imports and exports have brought rather interesting results – at first sight, no major problems were encountered during imports and exports; however, when the imported and exported data were checked, we found out that some of the data were kept and some were due to transformations performed by one of the citation managers lost.

The whole paper is focused on open source software – both institutional repository software and citation managers we have worked with are open source. The use of open source technologies has not been chosen arbitrarily – we believe that the importance of open source – as a parallel to Open Access movement – in library and information services belongs to current trends which should not be underestimated.⁴² On the other hand, we still have to keep in mind that open source does not mean perfect, however, it makes it possible for basically anyone to contribute with a piece of code. This may serve as an inspiration for (active) users institutional repository software (or its installation) which currently do not offer sophisticated citation export capabilities (unlike a number of mostly commercial digital libraries, fee-based journals or citation databases). For those institutions or individuals who consider setting up a repository, citation export capabilities may become one of selection criteria.

When users are concerned, the need to develop effective ways of citing documents available in various versions (say in an eprint archive and in a journal) becomes apparent. For example when creating references according to ČSN ISO 690 and especially ČSN ISO 690-2, a field for notes may be used to capture the relationship between document versions.

Last but not least we would like to point out that even complex citation formats (such as RIS) may cause problems when the fields are not used consistently by all systems. In our view further research into this topic would be useful for all stakeholders ranging from users (readers) to authors and publishers.

⁴² See for example: *Open Source Software in Libraries* or *LEE J. Open standards and software for bibliographies and cataloging*.

References

- A Guide to Institutional Repository Software* [online]. 3rd ed. Open Society Institute, August 2004 [cited 2009-05-12]. 28 p. Available from: <http://www.soros.org/openaccess/pdf/OSI_Guide_to_IR_Software_v3.pdf>.
- Comparison of reference management software. In *Wikipedia, the free encyclopedia* [online]. San Francisco (CA, USA) : Wikimedia Foundation, last modified on 10 May 2009, at 21:00 (UTC) [cited 2009-05-12]. Available from: <http://en.wikipedia.org/wiki/Comparison_of_reference_management_software>.
- ČSN ISO 690 (01 0197). *Dokumentace – Bibliografické citace – Obsah, forma a struktura [Documentation – Bibliographic references – Content, form and structure]*. Praha : Český normalizační institut, 1996. 31 p.
- ČSN ISO 690-2 (01 0197). *Informace a dokumentace – Bibliografické citace – Část 2: Elektronické dokumenty nebo jejich části [Information and documentation – Bibliographic references – Part 2: Electronic documents or parts thereof]*. Praha : Český normalizační institut, 2000. 22 p.
- CYZYK, Mark; CHOUDHURY, Sayeed. *A Survey and Evaluation of Open-Source Electronic Publishing Systems* [online]. Baltimore (MD, USA) : The Johns Hopkins University, April 28, 2008 [cited 2009-05-12]. 70 p. Available from: <<https://wiki.library.jhu.edu/download/attachments/22964/Open+Source+ePublishing+System+s+White+Paper.pdf?version=1>>.
- Institutional Repository. In *Wikipedia, the free encyclopedia* [online]. San Francisco (CA, USA) : Wikimedia Foundation, last modified on 21 April 2009, at 19:14 (UTC). [cited 2009-05-12]. Available from: <http://en.wikipedia.org/wiki/Institutional_repository>.
- KREJČÍŘ, Vlastimil. Systémy pro tvorbu digitálních knihoven [Systems for the digital libraries creation]. In *INFORUM 2006: 12. ročník konference o profesionálních informačních zdrojích, Praha 23.-25. května 2006* [online]. Praha : Albertina icome Praha, 2006 [cited 2009-05-12]. Available from: <<http://www.inforum.cz/archiv/inforum2006/prispevek.php-prispevek=24.htm>>. ISSN 1801-2213.
- LEE, John. 2007. Open standards and software for bibliographies and cataloging. In *SourceForge.net* [online]. April 2004 [cited 2009-05-12]. The list is no longer updated. Available from: <<http://wwwsearch.sourceforge.net/bib/openbib.html>>.
- LHOTÁK, Martin. Open source pro digitální knihovnu [Open source for a digital library]. In TKÁČIKOVÁ, Daniela; RAMAJZLOVÁ, Barbora (ed.). *Automatizace knihovnických procesů 2007 : sborník příspěvků* [online]. Praha : ČVUT, 2007 [cited 2009-05-12]. 8 p. Available from: <<http://www.akvs.cz/akp-2007/09-lhotak.pdf>>.
- Nature : international weekly journal of science* [online]. Nature Publishing Group, c2009 [cited 2009-05-12]. Weekly journal. Available from: <<http://www.nature.com/nature/index.html>>. ISSN 0028-0836 (print version). ISSN 1476-4687 (online version).
- Open Source Software in Libraries. *Bulletin of the American Society for Information Science and Technology* [online]. December/January 2009, vol. 35, no. 2 [cited 2009-05-12], p. 8-46. Special section of the issue. Available from: <http://www.asis.org/Bulletin/Dec-08/Bulletin_DecJan09_Final.pdf>. ISSN 1550-8366.
- RYGELOVÁ, Pavla. Zpráva z 2. setkání českých uživatelů systému DSpace [Report from the 2nd meeting of Czech DSpace users]. *Ikaros* [online]. 2009, roč. 13, č. 5 [cited 2009-05-12]. Available from: <<http://www.ikaros.cz/node/5431>>. URN-NBN:cz-ik5431. ISSN 1212-5075.
- Science : scientific research, news and career information* [online]. American Association for the Advancement of Science, c2009 [cited 2009-05-12]. Available from: <<http://www.sciencemag.org/>>.
- SKOLKOVÁ, Linda. Managing Citations and References with Software Tools. In *15th International Seminar CASLIN 2008 : Knowledge Technologies and (the Future of) Higher Education*. Brno : Masaryk University, 2008, p. 42-52. Presentation available from: <<http://alx.upol.cz/caslin2008/pom/skolkova.ppt>>. Videorecording available from: <<http://www.inflow.cz/caslin-2008-streamovane-video>>.

Website resources

- Connotea : free online reference management for all researchers, clinicians and scientists* [online]. Nature Publishing Group, c2005-2009 [cited 2009-05-12]. Available from: <<http://www.connotea.org/>>.
- CERN Document Server* [online]. Genève (Switzerland) : CERN, last updated: 12 květen 2009, 13:27 [cited 2009-05-12]. Available from: <<http://cdsweb.cern.ch/>>.
- CDS Invenio Demo* [online]. Genève (Switzerland) : CERN, last updated: 12-Dec-2008 [cited 2009-05-11]. Available from: <<http://cdsware.cern.ch/invenio/demo.html>>.
- CDS Invenio Overview* [online]. Genève (Switzerland) : CERN, c2005, last updated : 08 Aug 2005 14:41:47 MET [cited 2009-05-12]. Available from: <<http://cdsware.cern.ch/invenio/index.html>>.
- DSpace* [online]. Cambridge (MA, USA) : DSpace Foundation [cited 2009-05-12]. Available from: <<http://www.dspace.org/>>.
- DSpace* [online]. Cambridge (MA, USA) : DSpace Foundation [cited 2009-05-12]. Repository list. Available from: <<http://www.dspace.org/index.php/DSpace-Instances/Repository-List.html>>.
- E-LIS : E-prints in Library and Information Science* [online]. AEPI, [cited 2009-05-12]. Available from: <<http://eprints.rclis.org/>>.
- Fedora Commons* [online]. Ithaca (NY, USA) : Fedora Commons [cited 2009-05-12]. Available from: <<http://fedora-commons.org/>>.
- Fedora Commons* [online]. Ithaca (NY, USA) : Fedora Commons [cited 2009-08-26]. Fedora Commons Community Registry. Available from: <<https://fedora-commons.org/confluence/display/FEDINFO/Fedora+Commons+Community+Registry>>.
- ISI Web of Knowledge* [online]. Thomson Reuters, c2009 [cited 2009-05-12]. Available from (for authorized users only): <<http://isiknowledge.com/>>.
- Monash University ARROW Repository* [online]. Victoria (Australia) : Monash University, last updated: 20 Feb 2007 [cited 2009-05-12]. Available from: <<http://arrow.monash.edu.au/vital/access/manager/Index>>.
- Národní úložiště šedé literatury [National repository of grey literature]* [online]. Státní technická knihovna, last updated: 30 Jan 2009 [cited 2009-05-12]. Available from: <<http://nusl.stk.cz/index.php/Software>>.
- Open Access and Institutional Repositories with EPrints* [online]. Southampton (UK), University of Southampton [cited 2009-05-12]. Available from: <<http://www.eprints.org/>>.
- Open Access and Institutional Repositories with EPrints* [online]. Southampton (UK), University of Southampton [cited 2009-05-12]. Sites powered by EPrints. Available from: <<http://www.eprints.org/software/archives/>>.
- Registry of Open Access Repositories (ROAR)* [online]. Southampton (UK) : University of Southampton, c2005-2007 [cited 2009-05-12]. Available from: <<http://roar.eprints.org/>>.

ScienceDirect [online]. Amsterdam (Netherlands) : Elsevier, c2009 [cited 2009-05-12]. Available from: <<http://www.sciencedirect.com/>>.

Scopus [online]. Amsterdam (Netherlands) : Elsevier, c2009 [cited 2009-05-12]. Available from (for authorized users only): <<http://www.scopus.com/>>.

SpringerLink Home [online]. Berlin (Germany) : Springer, [cited 2009-05-12]. Available from: <<http://springerlink.com/home/main.mpx>>.

Welcome to DSpace at MIT [online]. Cambridge (MA, USA) : MIT Libraries [cited 2009-05-12]. Available from: <<http://dspace.mit.edu/>>.

Wiley InterScience [online]. Hoboken (NJ, USA) : Wiley, c1999-2009 [cited 2009-05-12]. Available from: <<http://www3.interscience.wiley.com/cgi-bin/home>>.

Zotero : The Next Generation Research Tool [online]. Washington, D.C. (USA) : George Mason University, Center for History and New Media [cited 2009-05-12]. Available from: <<http://www.zotero.org/>>.

VirtuReal Open Access 2.0

Jakub Štogr

`jakub.stogr@ff.cuni.cz`

Charles University in Prague, CZ, EU

Abstract: Open access can no longer be limited only to the providing access; such an initiative also needs to cover mechanisms of how to attract users of such sources and let them participate and involve them into the process of the added value creation and innovation. In order to create a sustainable and effective community around the open access resources, it is necessary to redefine the “approval mechanisms” of the scientific community, to boost social networking potential and to support the community-based solutions to minimize costs.

Keywords: open access 2.0, social networks, communities, attention, cooperation, innovation, interaction

Context of current Cha(lle)nges

“OA journals and repositories proliferated faster in 2008 than in any previous year. In 2008 the Directory of Open Access Journals grew by 812 peer-reviewed OA journals, or 27%. Last year it grew by 486 journals or 19%. In 2007, it added 1.4 titles per day, but in 2008 the rate jumped to 2.2 titles per day. It now lists a total of 3,812 peer-reviewed OA journals.”¹

The rapid growth in the numbers of journals and titles as described above – and proven by the evidence based on exact statistics from different sources² – can be seen worldwide. In both the USA and the EU we can find new initiatives that support the open access projects, some at the level of agreed policies, some coming with direct budget incentives.³

However, the rapid activity growth in the open access field can also be alarming. Thus, there are two general areas of study that need to be examined and possibly answered: (1) how this rapid growth affects the ability of users to pay attention and even differentiate between the content and its providers, and (2) whether these open access initiatives are based on realistic business models and whether they can become sustainable, cost effective and independent⁴. The initial setting of any open access platform that ignores these basic questions will not only influence the outcomes of the project itself, but it also influences the general environment of all open access initiatives.

Roots of Open Access

Open access was proclaimed in the Budapest Open Access Initiative in 2002 to peer-review the journal literature.⁵ The Initiative stated that although the peer-review process costs money, these costs are due to the existence of the Internet far lower than ever before; lower but there are some and these costs need to be somehow covered.

“Because price is a barrier to access, these new journals will not charge subscription or access fees, and will turn to other methods for covering their expenses. There are many alternative sources of funds for this purpose, including the 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, funds freed up by the demise or cancellation of journals charging traditional subscription or access fees, or even contributions from the researchers themselves. There is no need to favor one of these solutions over the others for all disciplines or nations, and no need to stop looking for other, creative alternatives.”⁶

In the “follow-up” initiative (the Bethesda Statement on the Open Access Publishing), we can find it necessary to ensure a long-term availability highlighted even more:

“...deposited ... in at least one online repository that is supported by an academic institution, scholarly society, government agency, or other well-established organization that seeks to enable open access, unrestricted distribution, interoperability, and long-term archiving.”⁷

¹ SUBER, Petr. *Open Access in 2008*.

² MORRISON, Heather. *Dramatic Growth of Open Access*.

³ SUBER, Petr. *Open Access in 2008*.

⁴ Both summarized in ESPOSITO, Joseph J. *Open Access 2.0*.

⁵ Budapest Open Access Initiative
<<http://www.soros.org/openaccess/read.shtml>>.

⁶ Budapest Open Access Initiative.

⁷ Bethesda Statement on Open Access Publishing.
<<http://www.earlham.edu/~peters/fos/bethesda.htm>>.

However, in the Berlin Declaration on the Open Access to Knowledge in the Sciences and Humanities⁸, we can find no peer-review explicitly mentioned – we can just read that:

“We define open access as a comprehensive source of human knowledge and cultural heritage that has been approved by the scientific community”.⁹

So, the very core idea itself comes with the paradox and “schizophrenic” situation when forcing the globalized world to take care of the open access movement as of the “public good”. This can be seen as a very crucial moment for the whole idea of open access together with the slightly shifted approach of the Berlin Declaration, stating a more detailed way of content „approval“ by the scientific community. This topic will be in depth investigated later.

Open Access 2.0

If we go back to the two previously mentioned questions and would like to further speak about the “open access 2.0”, we – as in the case of a phrase “web 2.0” – have to clearly state what precisely is meant by that.

As the Internet services evolve – based on changes that happened in the field of world wide web and how technology was used more interactively and user-centered – it is, at certain instances, becoming possible to see the differences in approaches, mainly in the sense of **users and their role**. If nothing more, this aspect together with the dot-com bubble burst can be marked as the birth of the web 2.0 era¹⁰. And the same logic can be applied to the issue of the open access when suffix “2.0” is being used.¹¹

For the following “2.0” mechanisms, we can use the “creative alternatives” mentioned in the Budapest Initiative as a possible solution to how to (not only) make the budget balanced:

- (1) expand what we call “scientific community” and/or redefine scientific community “approval” mechanism to allow wider participation;
- (2) strengthen community aspect of each repository, enrich “social networking aspects” inside these web-portals;
- (3) support open-source and community solutions to minimize costs for development, maintenance and up-to-date upgrades of software platforms;

⁸ Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities. <<http://oa.mpg.de/openaccess-berlin/berlindeclaration.html>>.

⁹ Berlin Declaration...

¹⁰ O'REILLY, Tim. *What Is Web 2.0*.

¹¹ Actually – as in the case of “web 2.0” – anyone has the right to simply declare his/her opinion that the “open access 2.0” is only a buzzword with no particular meaning.

Scientific community approach

Expanding scientific community can be seen as a radical approach. However, as we can assume by the increasing numbers of the open access journals and titles in them, this happens continuously. On one side, as the open access content is publicly available to anyone through the Internet, we can expect more people will find it “unintentionally”, e.g. as an answer to the basic search. On the other hand, as the blogosphere accepts and potentially very quickly disseminates any contribution in even shorter time than any “open access” mechanism, these inputs can influence wider public (and also members of the scientific community) and encourage people to get interested in science and its community.

Here we also touch the topic of what is the real reason for – not only scientific – publishing. We can highlight two basic aspects (1) publish to discuss and share (2) publish to self-present and get prestige. These aspects are basically not going against each other, and f.e. scientometrics has standard tools to measure “prestige”, but the overload at any of these aspects can damage the balance of the environment. Personal intentions are crucial element when assessing or trying to predict community potential; e.g. success of communication in “nautilus model” as presented by J. Esposito is based on possibility and ability to communicate among researchers freely (see below model *Figure 1*).

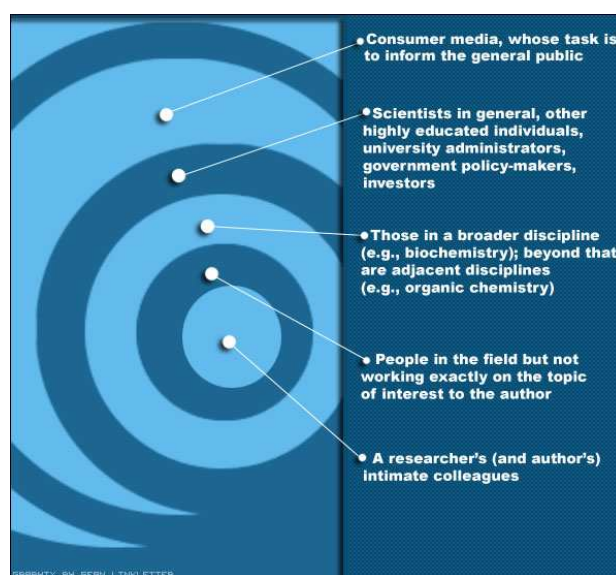


Figure 1 — Nautilus model for scholarly communications¹²

However, we can assume that following causal relationship cycles influence the scientific community:

- (1) the easier it is to establish the new journal (e.g. because of the wide Internet accessibility and easy ICT tools handling), the more journals will be established – there

¹² Taken from ESPOSITO, Joseph J. *Open Access 2.0*.

- are going to be more specialized journals and/or different level of quality settings of its content;
- (2) the more opportunities the scientists have to publish, the more they publish;
 - (3) the more cited articles from the open access repositories and journals exist, the more the researchers are going to publish in them; this also means that the more “resistant” and inflexible the traditional printed journals and publishing processes are – and the more supported the open access initiative is going to be – the more the scientists are going to publish under the open access mode.

Based on these assumptions, we can expect an exponential growth based on self-supporting mechanisms in the moment of reaching the critical mass – and based on statistics available right now, this moment is not far away from reaching, or – better to say – we already reached it.

Strengthening Community

In more details, strengthening community aspect does not mean limiting the functions of repositories to a sort of “virtual hard-drives”; the very aim should be interconnecting digital objects as well as connecting users – both authors and readers – into one transparent knowledge-based virtual network. This solution can be simply implemented using existing so-called “social network”¹³ API or through any other system – the logic is to simply enable and smooth the network interaction based on the scientific interests together with the personal profile and activities. Personal information and interactions have its special reason in these networks, mainly due to their unpredictability – once shared, the unique experience of an individual person becomes the attraction to other members of the community and can be afterwards used to start the communication “inside but outside” the scientific community; you can simply create (sub)networks of keen motorbike fans crossing over the community of psychologists and mathematicians discussing philosophy while meeting in a mountain pub.

As a commercial sphere example, we could mention the Amazon website, where you can find, among others, “also buy” hints, customer reviews, comments, etc. At a website like LinkedIn, you can directly share contacts, files, slideshows and other digital content, and regarding Facebook, probably the most powerful tool is the ability to create custom applications and implement them into the website itself.

Although it looks like strengthening the community around any particular repository or journal does not come with any direct cost benefits, the opposite could be the truth. Well described and active community with particular interest(s) can be simply targeted by any commercial

company, which can lead to generating a regular income from advertisements of sponsorship. An active community also updates the content of a website via comments of its members, which makes the websites not only more interesting, but also more SEO (search engine optimal /optimized).

Community sparing

The most significant and also simple solution how to lower the costs of open access is to use the community-driven and -based software (or even hardware) solutions. This solutions have actually already been implemented at some level. But not only the ICT resources can be shared. The whole community approach described briefly above can lead to more effective, distributed and collaboration-based solutions.

For example, a costly peer-review process can be substituted by a community-based quality evaluation through an optional “comment and weight” system. Such a “pre-review” can deliver almost the same feedback if implemented properly; moreover, through the identification of reviewers (if visible), the process can transparently show possible reasons for one's marking policy, e.g. the fact that s/he is working in a slightly different context if we use the “nautilus model” parallel.

Attention and Conclusions

“Open access 2.0” can be seen as a user-attention-centered approach due to the fact that any user is pushed and attracted by the networking environment to fully participate. All responses of users and interactions from the inside of the system are going to be disseminated widely through the network under seemingly unpredictable rules to boost innovation and fortuity. Nevertheless, all these activities and structures are driven by collaborative purposes and can be used to optimize processes and the usage of sources.

The unwelcomed heritage of the open access costs-acceptance can be substituted by the intentional engagement of social networks in their very original meaning as networks of personalities tied together with different kinds of (and not limited to) social linkages, experience and research interests. What more, it is necessary to interconnect with the idea of global responsibility to personally support open access initiatives by investing and donating our own time to lower the overall costs.

References

- Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities* [online]. Conference Open Access to Knowledge in the Sciences and Humanities. 20-22 Oct 2003, Berlin [cited 2009-05-13]. Available from: <<http://oa.mpg.de/openaccess-berlin/berlindeclaration.html>>.

¹³ Phrase “social network” is commonly used as an equivalent for the “Internet-based applications based on enhancement of social network aspects” – in this meaning it is also used in this article.

- Bethesda Statement on Open Access Publishing* [online]. 20 June 2003. [Cited 2009-05-13]. Available from: <<http://www.earlham.edu/~peters/fos/bethesda.htm>>.
- BJÖRK, Bo-Christer and HEDLUND, Turid. Two Scenarios for How Scholarly Publishers Could Change Their Business Model to Open Access. *Journal of Electronic Publishing* [online]. February 2009. vol. 12, no. 1. [cited 2009-05-13]. Available from: <<http://dx.doi.org/10.3998/3336451.0012.102>>.
- Budapest Open Access Initiative* (BOAI) [online]. 2002 [cited 2009-05-13]. Available from: <<http://www.soros.org/openaccess/read.shtml>>.
- ESPOSITO, Joseph J. Open Access 2.0: Access to Scholarly Publications Moves to a New Phase. *Journal of Electronic Publishing* [online]. Spring 2008, vol. 11, no. 2. [cited 2009-05-13]. Available from: <<http://dx.doi.org/10.3998/3336451.0011.203>>.
- HARLEY, Diane, EARL-NOVELL, Sarah, ARTER, Jennifer, LAWRENCE, Shannon and KING, C. Judson. The Influence of Academic Values on Scholarly Publication and Communication Practices. *Journal of Electronic Publishing* [online]. Spring 2007, vol. 10, no. 2. [cited 2009-05-13]. Available from: <<http://dx.doi.org/10.3998/3336451.0010.204>>.
- MORRISON, Heather. *Dramatic Growth of Open Access: Open Data Edition* [online]. 31 Dec 2008 [Cited 2009-05-13]. Available from: <<http://spreadsheets.google.com/ccc?key=pqCs8wrw32HHNKIMXUTgbSA&hl=en>>.
- MURKHERJEE, Bhaskar. Evaluating E-Contents Beyond Impact Factor - A Pilot Study Selected Open Access Journals In Library And Information Science. *Journal of Electronic Publishing* [online]. Spring 2007. vol. 10, no. 2. [cited 2009-05-13]. Available from: <<http://dx.doi.org/10.3998/3336451.0010.208>>.
- O'REILLY, Tim. *What Is Web 2.0 : Design Patterns and Business Models for the Next Generation of Software* [online]. 30 May 2005. [cited 2009-05-13]. Available from: <<http://oreilly.com/web2/archive/what-is-web-20.html>>.
- SUBER, Petr. Open Access in 2008. *Journal of Electronic Publishing* [online] 2009, vol. 12, no. 1. . [cited 2009-05-13]. Available from: <<http://dx.doi.org/10.3998/3336451.0012.104>>.
- SWAN, Alma Open Access and the progress of science. *American Scientist* [online]. 2007, 95 (3). . [cited 2009-05-13]. Available from: <http://eprints.ecs.soton.ac.uk/13860/1/American_Scientist_article.pdf>.
- WATERS, Donald. Open Access Publishing and the Emerging Infrastructure for 21st-Century Scholarship. *Journal of Electronic Publishing* [online]. 2008, vol. 11, no. 1. . [cited 2009-05-13]. Available from: <<http://dx.doi.org/10.3998/3336451.0011.106>>.

Open Access Institutional Repository and the Role of Academic Library Staff

Reconsidering the Position of the Academic Library in Preserving and Disseminating University's Scholarly Output

Daniela Tkačíková

daniela.tkacikova@vsb.cz

VŠB-Technical University of Ostrava, Czech Republic

Abstract: Libraries support for research has been most clearly identified by their collections of documents which are objects of research and collections of documents which report the results of research. Providing the technology like IR, and associated supporting services to access research output of the universities is a natural role for their libraries. The paper gives an overview of both present and planned activities of the VŠB-Technical University of Ostrava Central Library in setting up an efficient open access (OA) institutional repository (IR). The establishment of an institutional repository requires various issues to be considered and policy decisions to be made. In order to get more support from senior university management and academics the Library is planning to develop and recommend a set of guidelines concerning the content and submission policy, copyright issues, content policy etc. Establishing and running the university's IR by its library is part of the solution when thinking about the 'future' role of the university library.

Keywords: open access, institutional repositories, DSpace, scholarly communication, research support, university libraries, electronic theses and dissertations

Introduction

Research institutions, especially universities, in Europe and around the world have already set up their own institutional repositories (IRs) in order to make their research output openly accessible, to manage and disseminate as well as to preserve it. The IRs development and building has been inspired by the ideas of the Open Access (OA) movement. Thanks to this not only researchers from the institutions providing such services can benefit from the initiative to disseminate and make visible the research output more widely.

With an idea in mind that establishing an IR is a natural task for librarians, development of IRs has resulted in new responsibilities for academic librarians. To meet such demands, some university libraries created new types of positions, e.g. repository manager, intellectual property specialist or research data librarian, and the existed ones have been modified in order to help researchers in better understanding and appreciating the need for open access to scholarly literature.

I am strongly convinced that there is time to think about such ideas even in the Czech Republic. It is quite easy to find out (just using Google) that in the field of open access to scholarly publications institutions producing research publications in our country seems to be more likely behind the times.

Czech position on Open Access

As far as I know there is no official government policy or mandate to deposit research output arising from grants, no research funding agency requires depositing of publicly funded scholarly publications in the Czech Republic, nor exists any university which has made an institutional commitment to a mandatory policy for the deposit of published journal articles in its open repository in our country. This in fact also means there has been no active Czech response to the Recommendations from the European University Association (EUA) Working Group on Open Access¹ adopted by the EUA Council on 26th of March 2008 yet. The Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities² has been signed by the Czech Science Foundation – GACR (Grantová agentura České republiky) and also by the Academy of Sciences of the Czech Republic. The government of the Czech Republic can be found among the OECD Declaration on access to research data from

¹ EUA. *Recommendations from the EUA Working Group...* <http://www.eua.be/fileadmin/user_upload/files/Policy_Positions/Recommendations_Open_Access_adopted_by_the_EUA_Council_on_26th_of_March_2008_final.pdf>.

² Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities. <<http://oa.mpg.de/openaccess-berlin/berlindeclaration.html>>.

public funding³ signatories. Unfortunately this seems to be all that can be said about Czech research funding bodies' position on Open Access... And how many signatories from the Czech Republic have signed to the Petition for guaranteed public access to publicly-funded research results⁴? No wonder OA/IRs development in our country is still weak.

University libraries and OA/IRs

The minimum university libraries can both contribute and benefit from library staff 'know how' is their participation in promoting open access to scholarly output and engaging their universities in OA/IR movement. Libraries can also contribute to policy decisions and organizational issues around OA/IR development and show the ways in which their universities' research output can be made more visible and accessible by others without barriers. Establishing and running the university's IR by its library is part of the solution when thinking about 'future' role of university library. Lessons learned from abroad can help not only researchers, but also librarians in defining the library's policy relating to support and services for their university research activities. Moreover library can (or maybe better: should) act as a 'spiritus movens' of the process of adopting the ideas of OA at its university.

While librarians may be attracted to IRs for their potential and value as archives of institutional scholarly output, this alone is not likely to convince university management to undertake necessary administrative changes. In order to help the university to make an institutional commitment to open access to scholarly publications through the establishment of an institutional policy to deposit research output the libraries should develop a set of practices and procedures with regard to current needs of their university and its researchers. The content range and scope of a repository and future plans for the repository's development should all be taken into account too, as it is necessary to rely not only on an efficient software, technical infrastructure and IT staff support, but also on the library staff having the skills, and moreover time enough to provide such service for its university.

Setting up DSpace-based repository at VŠB-TUO Central Library

VŠB-Technical University of Ostrava (VŠB-TUO) Central Library⁵ had been interested in developing an institutional repository for some time and is already

engaged in establishing an OAI-compliant institutional repository. For many reasons the repository is still far from the ideas of OA movement ideas providing 'Closed Access rather than Open Access', unfortunately not using exactly the same sophisticated way as proposed by.⁶ But the first important step has been made as the present stage of the repository development and the content available can serve as a very important factor when demonstrating the benefits of an institutional OA/IR to the University.

It all started in 2005 when the Library made a decision to use open source software to provide access to information about a certain part of its holdings, i.e. VŠB-TUO theses and dissertations⁷, as well as to bibliographic data associated with scanned full-texts of articles published in the University's scientific journal *Transactions of the VŠB-Technical University of Ostrava*⁸.

As it is obvious in its initial phase the repository was not established with the intention of being 100% full-text and openly accessible to all its users. Until now most of the full-texts stored in the repository can be accessed by authorised users only, i.e. by students and staff of the University. Most of the repository content consists from bibliographic data only, esp. data associated to the Library collection of print theses and dissertations.

There are a number of software options available for setting up an IR. Different institutions have chosen different options in relation to the needs and functions that their IR will perform. At VŠB-TUO Central Library we wanted an open source solution, not only because we needed to build a repository at minimal cost, therefore DSpace⁹, the open source repository software developed by MIT and Hewlett-Packard was chosen.¹⁰ The repository software DSpace supports common standards for institutional repositories (XML, Dublin Core, OAI-PMH and METS) and it was chosen above EPrints¹¹ software, which has been also evaluated by the library staff as a possible solution. DSpace had better functionalities more suitable for the Library staff needs. It was easier to implement, maintain and use, especially from the point of view of our library with its lack of specialised IT staff and a low level support from the University IT Centre to this activity in the time of establishing the repository services.

⁶ HARNAD S. *How to Integrate University and Funder Open Access Mandates.*

HARNAD S. *Integrating University Thesis and Research Deposit Mandates.*

HITCHCOCK S. *Boost repository content with EPrints "Request eprint" button.*

⁷ [VŠB-TUO] *Kolekce vysokoškolských kvalifikačních prací ...* <<http://dspace.vsb.cz/handle/10084/14162>>.

⁸ [VŠB-TUO] *Sborník vědeckých prací...* <<http://dspace.vsb.cz/handle/10084/14242>>.

⁹ DSpace <<http://www.dspace.org/>>.

¹⁰ TKAČÍKOVÁ D. *Digitální knihovny – možnosti využití DSpace a EPrints.* TKAČÍKOVÁ D. *Zkušenosti s využitím DSpace v Ústřední knihovně*

¹¹ EPrints <<http://www.eprints.org/>>.

³ *Declaration on Access to Research Data from Public Funding* <<http://www.codata.org/archives/2005/UNESCOmtg/dryden-declaration.pdf>>.

⁴ *Petition for guaranteed public access to publicly-funded research results* <<http://www.ec-petition.eu/index.php?p=index>>.

⁵ *Ústřední knihovna VŠB-TUO* <<http://knihovna.vsb.cz/>>.

In order to set up its repository, the Library has adopted the DSpace version 1.3.2, and has installed it on a dedicated server. DSpace has proved to be a good choice in terms of the flexibility and functionality it offers. Maintaining the software has required minimal staff time. After the establishment of the technical infrastructure and the basic organisation within DSpace communities and collections and the Library staff accounts, the decision was taken to concentrate first on building up the repository content based on the data and associated full-texts of the collections mentioned above and assembled in the Library's internal databases. All the data and full-texts were successfully imported into DSpace environment and both librarians and users have started to use the living DSpace repository. The repository is designed to provide both data and services regarding primarily University theses and dissertations and research articles published by VŠB-TUO staff.

VŠB-TUO living DSpace installation can be found at <https://dspace.vsb.cz/>. Information from the repository can be retrieved through Google and Google Scholar, and the metadata have been also harvested by BASE¹² and OpenDOAR¹³.

One of the advantages of DSpace is its flexibility to organise the repository collections according to the University's faculties, institutes and departments structure. This also means that each department can have its own area on the repository if necessary, in order to give departments a sense of ownership over the metadata and deposited full-texts, and also that departments can link directly from their web pages to their staff publications on the repository, which seems to be a very important feature for promoting IR among the academics.

DSpace at VŠB-TUO: second stage

The second stage of the IR content range increasing has been inspired by the importance of information about the University's published research reported in Web of Science (WoS). Peer-review articles, impact factors and citations are of higher importance for the University's research evaluation and funding. Library staff has been involved in some activities around these issues, mainly in obtaining reliable data from WoS citation databases about publications of authors affiliated with VŠB-TUO. Searching for the relevant information has shown us that for such purposes it would be much better to build own database containing records of articles published in Impact Factor (IF) journals as the source data in WoS do not contain reliable affiliation information.

In order to be able to provide comprehensive information about the University's research publications recorded in WoS the Library decided to use DSpace repository as a container of metadata for University's IF

publications. Therefore a new collection¹⁴ based on information from WoS and containing records of articles published from 1990 to present has been created.

Even though this collection provides bibliographic information about the most important University's research publication only it serves as a sort of showcase of the University's intellectual quality and can be used for some evaluation purposes, and obviously not only for them. In addition abstracts and links to the full-texts at publishers' sites (or at subject and other open repositories) have been provided if available.

Soon after the collection's introduction to the users the Library has received positive feedback from the academics and the management of the University. This collection has been a good starting point for the future repository development. It is obvious it will be used as a basis for the planned repository enhancement and its change into the real repository containing full-texts of research publications.

It is also important to note that the Library has used a useful DSpace feature to map the items from the collection with the department communities and also with a collection based on IF journals¹⁵ only. This option to map items across more than one collection has been also useful for the articles by researchers from different departments working together on a research project.

Although VŠB-TUO Central Library had established its repository in 2006 for the potential purposes of showcasing the full-texts of research output of the University, as there are no official governmental, funding agencies or other institution, nor the University's requirements to provide open access to published research, there are also no depositing activities in the repository until now. Setting up a repository is one thing, but getting users to participate in its ongoing development is quite another one. The experience is when nobody requires depositing the refereed publications based on public funding in OA/IR, nobody is going to deposit them as the researchers are still satisfied enough with having published the article in an IF journal. It is also well-known fact the researchers' uptake has been reluctant where voluntary compliance is needed.

2009: Electronic theses 'pilot'

The present situation at our university in the field of its repository building shows it is time to reinvent all that has been done so far. As mentioned above the VŠB-TUO DSpace repository is still not OA/IR, nor it is the repository of full-texts even though there is a large amount of full-texts in it now because of the last academic year's 'submission' of three faculties' theses¹⁶. This has been

¹² BASE <<http://digital.ub.uni-bielefeld.de/index.php?l=en>> 39 951 items from <<http://dspace.vsb.cz/>> have been harvested until now.

¹³ OpenDOAR <<http://www.opendoar.org/>>.

¹⁴ [VŠB-TUO] *Publikační činnost akademických pracovníků VŠB-TUO* <<http://dspace.vsb.cz/handle/10084/56138>>.

¹⁵ [VŠB-TUO] *Publikační činnost v časopisech s IF podle JCR* <<http://dspace.vsb.cz/handle/10084/65861>>.

¹⁶ [VŠB-TUO] *Plné texty bakalářských a diplomových prací...* <<http://dspace.vsb.cz/handle/10084/65349>>.

a successful test of the DSpace repository. Thanks to it the University management has made a decision about using the repository as a space for a long term preservation and access to VŠB-TUO theses. This has proven the Library's effort and work undertaken in recent few years.

As mentioned above theses are not openly available, they can be accessed by authorised users only, as there is no University's policy on open access, and even more, there is no policy of mandatory submission because of some issues that have not been solved yet, such as copyright, fear of plagiarism, etc. But from the Library point of view the most important decision has been made. Although still only on a voluntary basis, the students have been asked and encouraged to submit electronic versions of their theses into the University Information System, called Edison, which has been designed and enhanced in order to allow PDF files (not only) of theses submission. Soon after final exams results in the end of the academic year all e-theses PDF files uploaded, and of course associated metadata too, will be imported into DSpace database.

The process is described simply although a lot of work is still expected. There are a number of technical and other issues that still need to be solved, e.g. intellectual property (the Library will suggest the University to use Creative Commons 'Attribution-Non Commercial-Share Alike' licence in order to simplify future rights and permission process), but the Library staff believe this academic year experience will result in a mandatory deposit of electronic theses at our university starting next academic year, at least at some faculties.

Until now no decision relating dissertations (PhD theses) has been made. It is also important to emphasize that the Library do not have any plans to digitise its retrospective collection of print theses and dissertations. All activities around electronic versions of theses concern 'digitally born' theses only.

DSpace at VŠB-TUO: next stage

It is the primary challenge for the Library now to prepare a next stage of the repository active use by the researchers and deposit published peer-reviewed papers in order to help adoption of the OA/IR by the University. The activities planned may possibly result in an institutional commitment to an open access repository and a mandatory policy for the deposit of published (peer-reviewed) journal articles.

The establishment of an institutional repository requires various issues to be considered and policy decisions to be made. In order to get more support from the senior University management and academics the Library is planning to develop and recommend a set of documents concerning the content and submission policy, copyright issues, metadata and preservation policy, etc. The most important thing is to establish positive relations with the content contributors, researchers. Before they can contribute, the researchers must know that such repository

exists. Studies and surveys conducted and published recently have found that academics lack awareness of the existence of IRs, even in countries where the ideas of the OA movement have been widely adopted by universities and other research institutions and where the support for the OA has been accompanied not only by a growing number of repositories, but also by the research funders' policies requiring the researchers to deposit published research output.¹⁷

Therefore the primary focus of the Library will be to increase understanding, awareness, recognition, and visibility for the OA/IR within the University, as well as explaining OA movement ideas to its senior management and research staff. It has been one of the key lessons learned from abroad that promoting is crucial for OA/IR successful implementation. The most important strategic task for the Library in the near future will be focused on advocacy and promoting to encourage the researchers to put their papers to the repository.

In order to achieve all its aims the Library is going to prepare a lot of necessary activities which include areas such as:

- Promoting;
- Advocacy;
- Full-text content recruitment;
- Advice on copyright, licensing and other intellectual property issues;
- Content policy, etc.

Library website and a new blog, presentations and other promotional activities will be prepared in order to present comprehensive information of potential benefits of open access to scholarly publications. It will be also useful to publicise and promote the repository and the OA movement ideas through the VŠB-TUO magazine *Akademik*¹⁸.

It is very important to have a good communication strategy to overcome unawareness, doubts and reluctance from academics. As the researchers seek for citations they have to know the more research articles will be open it is more likely they will be read and possibly more cited. There is a growing body of papers showing that the impact of research is substantially enhanced by deposit in an OA/IR.¹⁹ It is also necessary to convince the authors that the benefit from OA should be reciprocal. They cannot be users of OA documents only; they should be producers as well.

¹⁷ DAVIS P.M.; CONNOLLY M. J.L. *Institutional Repositories*

FOSTER N.F.; GIBBONS S. *Understanding Faculty to Improve Content Recruitment for Institutional Repositories*

¹⁸ *Akademik* <<http://www.vsb.cz/okruhy/nastroje-sluzby/infoservis/akademik>>.

¹⁹ HAJJEM, C.; HARNAD, S.; GINGRAS, Y. *Ten-year cross-disciplinary comparison of the growth of Open Access and how it increases research citation impact.*

See also

The effect of open access and downloads ('hits') on citation impact: a bibliography of studies, <<http://opcit.eprints.org/oacitation-biblio.html>>.

Concerning the full-text content policy the repository will contain only material that has already been published through the established scholarly channels, i.e. peer-reviewed articles, in the repository content building initial phase journal articles only. The best way is to use the DSpace collection of metadata for IF journals articles published by VŠB-TUO academic staff.

It would be possibly easier to attract senior VŠB-TUO researchers, mainly those responsible for leading research projects, who are also more closely involved in publication activities, and more using library services, therefore their approach to the Library activities would be more positive in general. Thanks to the above mentioned DSpace collection of articles metadata the Library will be able to select three or maybe more most important authors from each faculty or whole-university institute to contact them and ask for collaboration on depositing of full-texts of their papers into repository. Such 'early adopters' and their full-texts can act as effective marketing agents for the repository.

Unfortunately we cannot expect the content automatically flowing into the repository through the process of 'self-archiving' by authors. It is a well-known fact that researchers do not like the time consuming work to spend on – from their point of view – administrative issues, and they use other staff or student assistants to undertake the citation checking and other necessary activities relating to reporting the research output. Understanding behaviours and attitudes of researchers as well as not expecting the whole-of-university mandate for the deposit of its research outputs, the Library policy is not to insist on self-archiving.

As part of its strategy to engage the academics in their full-texts deposit, the Library staff is going to do the copyright checking on behalf of authors. In order to get some initial content it would be necessary to identify some so-called 'green' publishers who allow self-archiving in any form. This will be done using the SHERPA/RoMEO²⁰ database. Then we will ask the authors who have published their papers in those journals for permission to deposit the papers in our DSpace IR.

We expect that library mediated submission will be welcomed by academics as it will only need submitter to pass to the Library an electronic copy of the article to be placed in the repository. Mediated deposit service will be also useful to monitor the content of the repository, as well as its quality. Such deposition should not, however, be a long-term solution as the best scenario is to get authors to perform their own self-archiving.

Although the ideal option would appear to be a formal requirement for researchers to place their research output in the repository, accompanied by effective author support policies to start checking publisher permissions, the decision was made by the Library not to push for formal policy and mandate. The surveys and studies show that steady growth of IR content can be found everywhere the

universities have made an administrative investment requiring deposits of their research publications. The necessary mandate enforcement will need to be negotiated with senior research management but before all this critical mass would be achieved in terms of repository content.

With all this in mind and knowing that academics are rather unwilling to put time and effort to new activities requiring some level of IT literacy and some other skills the Library knows the way to the sustainable OA/IR will be long and not easy.

To promote and facilitate possible future self-archiving by researchers in OA/IR it would be better to assist researchers, at least during the initial period of obtaining peer-reviewed full-text content for the repository. This would also be a good opportunity to show that librarians can serve as essential human infrastructure in OA/IR development and help the University to use its repository as an efficient management tool for its research management.

Conclusion

Czech university libraries are understaffed and still focused mainly on traditional services which support educational activities of their institutions more than their research. But I strongly believe that despite of the situation it is very important to reconsider university libraries' activities right now if the libraries do not want to fall behind. Libraries support for research has been most clearly identified by their collections of documents which are objects of research and collections of materials which report the results of research. Providing the technology like IR, and associated supporting services to access research output of the universities is a natural role for the libraries. University libraries must be able to respond to the changes in scholarly communication otherwise they risk becoming irrelevant as partners in their institutions research process.²¹

References

- DAVIS, Philip M.; CONNOLY, Matthew J. L. Institutional Repositories: evaluating the reasons for non-use of Cornell University's installation of DSpace. *D-Lib Magazine* [online]. Mar 2007, vol. 13, no. 3/4 [cited 2009-05-12]. Available from: <<http://dx.doi.org/10.1045/march2007-davis>>. ISSN 1082-9873.
- FOSTER, Nancy Fried; GIBBONS, Susan. Understanding Faculty to Improve Content Recruitment for Institutional Repositories. *D-Lib Magazine* [online]. Jan 2005, vol. 11, no. 1 [cited 2009-05-12]. Available from: <<http://dx.doi.org/10.1045/january2005-foster>>. ISSN 1082-9873.
- HAIJEM, C.; HARNAD, S.; GINGRAS, Y. Ten-year cross-disciplinary comparison of the growth of Open Access and how it increases research citation impact. *IEEE data engineering bulletin*. 2005, vol. 28, no. 4, p. 39-47. Available also from: <<http://eprints.ecs.soton.ac.uk/12906/>>.

²⁰ *Sherpa/RoMEO. Publisher copyright policies & self-archiving.* <<http://www.sherpa.ac.uk/romeo>>.

²¹ *WAINWRIGHT E. The future of the 'research' library in an age of information abundance and lifelong learning.*

- HARNAD, Stevan. *How to Integrate University and Funder Open Access Mandates* [online]. March 2, 2008 [cited 2009-05-12]. Available from: <<http://openaccess.eprints.org/index.php?/archives/369-guid.html>>.
- HARNAD, Stevan. *Integrating University Thesis and Research Deposit Mandates*. Open Access Keynote Address at ETD 2009 [cited 2009-05-12]. Available from: <<http://conferences.library.pitt.edu/ocs/viewpaper.php?id=750&cf=7>>.
- HITCHCOCK, Steve. *Boost repository content with EPrints "Request eprint" button* [online]. 7th April 2006. [Cited 2009-05-12]. Available from: <http://www.eprints.org/news/features/request_button.php>.
- TKAČÍKOVÁ, Daniela. Digitální knihovny – možnosti využití DSpace a EPrints. In MIKULECKÁ, Jaroslava; SEDLÁČEK, Jan (sest.). *Digitální knihovny: provoz a budování : sborník konference EUNIS-CZ : Špindlerův Mlýn, 28. až 30. května 2006*. Hradec Králové : Gaudeamus, 2006, p. 103-108. ISBN 80-7041-200-3.
- TKAČÍKOVÁ, Daniela. Zkušenosti s využitím DSpace v Ústřední knihovně VŠB-TU Ostrava. In *Infos 2007 : 34. mezinárodní informatické sympóziium : 16.-19. apríla 2007, Stará Lesná* [CD-ROM]. Bratislava : Spolok slovenských knihovníkov, 2007. ISBN 978-80-969674-0-7.
- WAINWRIGHT, Eric. The future of the 'research' library in an age of information abundance and lifelong learning. *Australian Academic & Research Libraries* [online]. September 2005, vol. 36, no. 3 [cited 2009-05-12]. Available from: <<http://www.alia.org.au/publishing/aarl/36.3/full.text/wainwright.html>>. ISSN 0004-8623.
- Website resources**
- Akademik : časopis VŠB - TU Ostrava* [online]. Ostrava : Infocentrum VŠB-TUO, 2001- <<http://www.vsb.cz/okruhy/nastroje-sluzby/infoservis/akademik>>. ISSN 1213-8916.
- BASE (Bielefeld Academic Search Engine) [online]. Basic Search. [Cited 2009-05-12]. Available from: <<http://digital.ub.uni-bielefeld.de/index.php?l=en>>.
- Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities* [online]. Conference Open Access to Knowledge in the Sciences and Humanities. 20-22 Oct 2003, Berlin [cited 2009-05-12]. Available from: <<http://oa.mpg.de/openaccess-berlin/berlindeclaration.html>>.
- Declaration on Access to Research Data from Public Funding* [online]. Paris, 30 Jan 2004 [cited 2009-05-12]. Available from: <<http://www.codata.org/archives/2005/UNESCOmtg/dryden-declaration.pdf>>.
- DSpace [online]. Open source software. [Cited 2009-05-12]. Available from: <<http://www.dspace.org/>>.
- EPrints [online]. [Cited 2009-05-12]. Available from: <<http://www.eprints.org/>>.
- EUA (European University Association) *Recommendations from the EUA Working Group on Open Access adopted by the EUA Council on 26th of March 2008* [online]. University of Barcelona, Spain [cited 2009-05-12]. Available from: <http://www.eua.be/fileadmin/user_upload/files/Policy_Positions/Recommendations_Open_Access_adopted_by_the_EUA_Council_on_26th_of_March_2008_final.pdf>.
- OpenDOAR (Directory of Open Access Repositories) [online]. [Cited 2009-05-12]. Available from: <<http://www.opendoar.org/>>.
- Petition for guaranteed public access to publicly-funded research results [online]. [Cited 2009-05-12]. Available from: <<http://www.ec-petition.eu/index.php?p=index>>.
- Sherpa/RoMEO. Publisher copyright policies & self-archiving [online]. [Cited 2009-05-12]. Available from: <<http://www.sherpa.ac.uk/romeo>>.
- Ústřední knihovna VŠB-TUO [VSB-Technical University of Ostrava. Central Library] [online]. [Cited 2009-05-12]. Available from: <<http://knihovna.vsb.cz/>>.
- [VŠB-TUO] Kolekce vysokoškolských kvalifikačních prací zpracovaných do konce 1. pololetí 2006 : [20904] [online]. [Cited 2009-05-12]. Available from: <<http://dspace.vsb.cz/handle/10084/14162>>.
- [VŠB-TUO] Plné texty bakalářských a diplomových prací HGF, FBI, EKf (export do Národního registru VŠKP r. 2008) : [830] [online]. [Cited 2009-05-12]. Available from: <<http://dspace.vsb.cz/handle/10084/65349>>.
- [VŠB-TUO] Publikační činnost akademických pracovníků VŠB-TUO : [892] [online]. [Cited 2009-05-12]. Available from: <<http://dspace.vsb.cz/handle/10084/56138>>.
- [VŠB-TUO] Publikační činnost v časopisech s IF podle JCR [online]. [Cited 2009-05-12]. Available from: <<http://dspace.vsb.cz/handle/10084/65861>>.
- [VŠB-TUO] Sborník vědeckých prací Vysoké školy báňské - Technické univerzity Ostrava : [4268] [online]. [Cited 2009-05-12]. Available from: <<http://dspace.vsb.cz/handle/10084/14242>>.

Open Access and Institutional Repository

Approach of the Czech Academy of Sciences and Link-up to International Projects

Martin Lhoták & Roman Chýla

Lhotak@knav.cz, Roman.Chyla@gmail.com

Czech Academy of Sciences Library, Czech Republic

Abstract: The Academy of Sciences of the Czech Republic (ASCR) is a signatory of the Berlin Declaration since the last year. In April 2009 the Scientific Council of the Academy stepped forward and recommended in a statement addressed to the newly elected management of the Academy to: formulate ASCR policy for OA; finish preparation of the central institutional repository of ASCR and use it for OA; actively use ASCR's potential as a unified and strong organization for effective development of an information infrastructure for R&D and determine concrete rules for access to infrastructures built in the frame of "OP VaVpl EU" projects.

The Czech Academy of Sciences Library (CASL) has been invited to participate in the process and actively contribute to the OA case in the future. There are a few projects with central impact that concerns OA. We describe two of them in the paper.

Keywords: open access, intitutional repositories, digital library, projects, scholarly publishing, publisher

Approach to Open Access

The Academy of Sciences of the Czech Republic (ASCR) is the biggest non-university research institution in the Czech Republic with more than 100 year history. ASCR consists of more than 50 autonomous institutes, each specialized into specific field, annually producing more than 11 000 scientific papers. Many of institutes also run their own scientific journals and the Academy, as a whole, own a publishing house Academia.

Even if ASCR is a signatory of the Berlin Declaration since the last year – and the very first signatory from among all institutions in Czech Republic – the appeal of Open Access movement has not been tangible on the top management level until very recently. There were only few signals that a consistent approach will be adopted by the council. The Grant agency of ASCR (as well as any other funding body in Czech Republic) does not require any open-access provision, no mandate. There exist no clear plans to create a central repository. The fact that each institute at ASCR is an independent unit does not help in this respect. Any centralized approach is collectively resisted unless there are economical incentives, success of any Open Access (OA) initiative thus depends very much on the individuals and their personal undertakings from inside and outside the ASCR.

But recently – in April 2009, - the Scientific Council of the Academy stepped forward and recommended a formulation of a policy to the newly elected management of the Academy. Scientific Council does not belong to an executive branch of the ASCR, nevertheless

it is a powerful part with advisory influence. It was recommended to:

- a. Formulate ASCR policy for OA
- b. Lay down rules for a central institutional repository of ASCR and use it for OA
- c. Actively use ASCR's potential of a unified and strong organization and develop information infrastructure for R&D
- d. Determine concrete rules for access to infrastructures that was built using resources from EU projects

The first three points are politically significant, because they show that scientists are acutely aware of the OA provisions and they wish their mother organization to adopt a clear stance – and not only adopt, but possibly enforce it. The Czech Academy of Sciences Library (CASL) has been invited to participate in the process and actively contribute to the OA case in the future. The hopes that in a relatively near future ASCR will adopt one policy are a bit higher, the situation is changing and it is possible that more active approach awaits us.

Yet even if there are many barriers, economic concerns and political issues, we cannot say the before mentioned activity arrived all of sudden. What looks quiet from above will be seen very much differently from the bottom-up perspective. There are a few projects with central impact that concerns OA. We would like to describe two of them here, but many other OA projects inside institutes has already started to have an impact on minds of editorial boards members – at least in fields where advancement of science is measured in years rather than in decades.

Library as “a publisher”

The title of this section is perhaps too bold, but nevertheless it clearly states the direction where CASL is heading with regard to electronic publishing. Inside the ASCR there exist many scientific journals, but because they are independent, also their work habits, workflows and policies are different. But when we tried to map the situation, what was common, surprisingly, was the level of “dis-advancement” of the editorial work. Journals are publishing quality scientific publications, but they do so still using very basic tools and methods – it is not very far from reality to note here paper notebooks, calendars and e-mails. Whatever is beyond this is also beyond level of the tech-experience of the executive editors. And this situation does not only have implication for them, it negatively bears on the attractiveness of the journals for other authors, reviewers and sometimes also readers. It impairs budget of the journal and its ability of adaptation to new trends. This mind configuration is perhaps best described as “fierce resistance to anything new”. It is affirmed by our experience and also the survey we conducted in early 2008.

The survey, in which 39 scientific journals participated, showed an interesting distribution of the production. Almost half of the journals respondents managed annually less than 50 manuscripts. Another 9 journals were also of local importance but slightly bigger, with their average reviewed manuscripts reaching 80 manuscripts. Finally the last group was made of 10 journals with an average of 130 reviewed manuscripts. And quite significantly, this quarter of the journals was responsible for almost 50% of all the published content. Worth mentioning is the fact that there exists great differences between journals in terms of technology they use, publishing practices, expectations, amount of work, human resources and also budget. If we were to address the situation for all the ASCR journals, the project would be doomed from the onset.

If two quarters has reacted with resistance to a new technology advancement, the third quarter has hesitated, the last quarter is everything but different. Journals in this area represent internationally influential publications and the fact that the production is prepared “in the old way” has an adverse effect on them – there is a strong competition and they do mind if somebody gives them competitive advantage. These journals were more than willing to – and to be fair, the impulse for change came from one of the journals. Therefore CASL jointly with experts from journals editors compared available systems, prepared project and obtained resources from the central budget (credit should go also to the ASCR management for supporting the initiative financially).

Journals, with their own resources could not even contemplate similar undertaking alone and the same applies to their situation and OA. They cannot seriously start considering going open-access (even if some of them would like to) without reducing cost and at the same time attracting more authors to publish in their publications. In this CASL has become their partner in the publishing

process and the services CASL provide and plans to provide are an important contribution to the publishing.

We are aware of a need to gather more data, more evidence of the impact, but currently there are a few plausible scenarios:

Publishers (like Springer) will bid even more than now and journals will no longer resist the temptation and “sell themselves” to the big publisher houses. This is slightly contradicted by the stronger position the journals acquire. At least they may have different position for negotiation.

Institutes will contemplate making the job of publishing profitable (again, after some time). Given the nature of the publishing and the fact that profit is not and was not the reason of the existence, our journals may continue to exist and be subsidized as many other journals in the world.

Some journals will decide to go open-access, but probably not before they have the publishing infrastructure at place and the cost of running the journal does not slip too much into the red numbers. The journals are and probably will continue to be dependent on the institute’s budget, but open-access must not worsen this situation. Only if this problem is solved in a satisfactory manner can journal make the radical change from the traditional into an open-access mode.

Certainly, the situation is not as simple as the three options say, but whatever the future, it is certain that most scientific journals of the ASCR are not going to be fully open-access in a near future. There are cases of some that are OA, but also the case of one that has tried and failed, and is now in a vegetative state between life and death. CASL will nevertheless continue to support the journals and take some burden from their shoulders, provide services and take care of the solutions that can help to make the transition to open-access possible. One of the solutions is the ongoing project of a digital library.

Digital library and institutional repository

Results from another survey conducted by ASCR among the Academy institutes in 2008 shows interesting numbers regarding open access publishing. 13 out of 46 journals would agree to make the recent production freely available, 29 of them would open the historic issues digitized by CASL (this archive is currently available only on the premises of CASL and contains circa 1. million pages on-line). But the most pressing issue, and that is evident from the discussions with journal managers, is again money – journals are concerned about their position after the production is made available online and they usually adopt very conservative stance and wait. A lot of work is due in the field of marketing and persuasion.

The repository that CASL is building has attracted attention of the 15 journals so far. Two of them are completely OA (green OA) and with the help of the digital library, their production is disseminated and propagated to many different resources. We are trying to convince other institutes to join the effort. CASL is

working on agreements to provide the old and eventually also the new content online and after the initial hesitation, the project seems to be taking off.

The situation is complicated by decisions on multiple levels. Sometimes, the management of an institute is in favor of electronic publishing, or only some members of the journal or it can be a combination of the above and sometimes the effort breaks down simply because just the administrative workers refuse to learn anything new. Not surprisingly – a new technology brings along a new challenge and we have to be careful to make the transition as simple and easy as possible. People value “word of mouth” and listen to experiences of their colleagues, especially if those colleagues are from the same institute, and slowly the idea of a digital library is getting better position among the journals.

CASL plans to enhance the current version with many new features, especially with pay-per-view and other subscription tools to actually give the journals means to sell the production. It is our experience that journals appreciate the effort somebody is investing in them and helping them. And if it is not contrary to their existential interests, they will contemplate giving back – and in this case, it is the role and goal of the CASL to build a complete archive of the ASCR scientific output and if possible, create conditions for open-access publishing. Therefore if journals allow, we will try to bring more of the scientific content to scientific community around the world.

A different case is an institutional repository of the ASCR. CASL will probably have to concentrate its activities on the management – it is mainly a political decision to foster OA inside the Academy and even if the technical solution is necessary, CASL has experience and expertise in providing the solution. Yet the solution without ASCR commitment will only mean wasted resources. Our hopes are in the current changes as described at the beginning of the article. The Scientific Council has made clear recommendations to ASCR management and CASL is involved and may try to provide further incentives and drive the effort. CASL can offer long term preservation of ASCR e-material, which is

another plus for institutional repository that should be advertised. A completely new solution for the long term preservation will be implemented in 2009.

Link-up to International Projects

As far as we are able to provide more a more material openly accessible, we have started to look around, trying to find eligible EU project(s) to join.

Generally encouraging is approach of EU officials to OA that states: “The Commission is currently experimenting with open access by making results of the EU funded research accessible to all online.”

There are 2 main EU funding platforms for OA:

- FP7 – Seventh framework programme for research and technological development in OA – € 75 mil. in 2007-2008
- EU program CIP ICT PSP (CIP – Competitiveness & Innovation Framework Programme)

Theme 2: Digital Libraries € 25 mil. in 2009

Objective 2.4: Open access to scientific information

For the start we decided to apply for participation in the DRIVER II project, which primary objective is to create a pan-European infrastructure for digital repositories, offering sophisticated services and functionalities for researchers, administrators and the general public. In DRIVER search portal is today approximately 1 000 000 documents (journal articles, dissertations, books, lectures, reports, etc.) harvested regularly from more than 200 institutional or thematic repositories from 23 European countries in 25 languages.

Another EU project, where we have started to communicate about our collaboration is the Europeana, which is also oriented only to openly accessible material. We should join the Europeana v1.0 Thematic Network in the very near future.

Better interoperability, visibility and wider dissemination of ASCR publications are goals, which we believe we could achieve by joining these projects.

An Experience with Building Digital Open Access Repository DML-CZ

Petr Sojka

sojka@fi.muni.cz

Faculty of Informatics, Masaryk University Brno Czech Republic

Abstract: A successfully built institutional or community repository (e.g. set of workflows) needs a coordinated effort of librarians, IT specialists and representatives of users – content specialists. We will explain and discuss design, technical and political decisions behind building the Czech Digital Mathematics Library (DML-CZ <http://dml.cz>) in the context of other successful thematic community projects (PubMed Central, ADS, SCOAP3 and planned EuDML). A framework developed for handling different types of mathematical publications is presented. It integrates workflow for the articles scanned from a paper, for documents from retro-born digital period (data were available in some type of electronic form) and for born-digital papers (newly published data from publishers). Experience gained, lessons learned and tools prepared during development of the Czech Digital Mathematics Library DML-CZ are presented. We describe problems of migration of existing workflows (born-digital, retro-digital) into the repository, negotiations with Google Scholar towards better visibility, indexing and search, and problems of copyright and sustainability issues we have faced.

Keywords: digital mathematical library, mathematical knowledge representation, workflow, retrodigitization, DML-CZ

Motivation

Digital Library (DL) business has moved from data/files centered processing towards process-oriented *workflows*. Workflows enact the machinery of building and running a digital library. Instead of mirroring file repositories more subtle solutions have to be devised: data curatorship changes to workflow curatorship and *services*.

World and its DLs are becoming global. Some methods (e.g. citation ranking) start to work only as part of global, world-wide system. On the other hand, focus of search in [global] DLs inevitable has to support narrowing and semantic filtering for the needs of specific communities. It is the case also for the mother of sciences, *mathematics*.

There are communities and systems that start to dominate in some thematic areas: PubMed Central (PMC)¹ is one of such system in medical domain, speeding up the research and author's citation indexes in the area. Unfortunately, only domains where global initial funding was available took advantages of the platforms established and tools and workflows were developed and reached the critical mass. In the PMC case, journal publishers are now eager to join the club, authors enjoy global topical ontology-based search. Researchers prefer to send their papers to journals available in PMC as this leverages their citation indexes.

Domain of mathematical publications has not yet reached similar stage, although there are referative

databases like Mathematical Reviews² or Zbl³. These databases contain additional independent reviews, but they do not have full texts of articles and thus losing most of today's possibilities stemming from full-text availability and search. Another problem with mathematical publications is that they often contain many formulae that are hard to optically recognize and standard DL systems do not support their proper handling on [full-]text level. So far no significant initial funding for Worldwide Mathematical DL (WMDL) was successful, leaving the floor open for 'bottom-up' smaller initiatives and projects as NUMDAM/ CEDRAM/ CEDRICS⁴, Euclid, Jahrbuch, RusDML⁵, arXiv or Czech Digital Mathematics Library (DML-CZ)⁶

In this paper we describe a framework developed for handling different types of mathematical publications in the

² *Mathematical Reviews*. <<http://www.ams.org/mr-database>>.

³ *Zentralblatt MATH* <<http://www.zentralblatt-math.org/zmath/>>.

⁴ *NUMDAM* <<http://www.numdam.org/>>.

CEDRAM <<http://www.cedram.org/>>.

BOUCHE T. Next Digital Mathematics Library.

⁵ *Euclid* <<http://projecteuclid.org/>>.

ERAM. The Jahrbuch Project. <<http://www.emis.de/projects/JFM/>>.

Math-Net.Ru. <<http://www.mathnet.ru/>>.

WEGNER B. RusDML 2008: Current Facilities of the Core Archive of Digitized Russian Publications in Mathematics.

⁶ *arXiv* <<http://arxiv.org/>>.

DML-CZ <<http://dml.cz/>>.

SOJKA P. From Scanned Image to Knowledge Sharing.

BARTOŠEK M., LHOTÁK M., RÁKOSNÍK J., SOJKA P. and ŠÁRFY M.

DML-CZ: The Objectives and the First Steps.

¹ *PMC* <<http://www.pubmedcentral.nih.gov/>>.

project DML-CZ. It integrates workflow for the articles scanned from a paper, for documents from retroborn digital period (data were available in some type of electronic form) and born-digital (newly published data from publishers). We report on the experience gained, lessons learned and tools prepared during the development of the DML-CZ project.⁷

The aim of the project approved for the five years period 2005–2009 is to digitize the relevant mathematical literature published in the Czech lands. It comprises periodicals, selected monographs and conference proceedings from the nineteenth century up until currently produced mathematical publications. It has been launched and is available on dml.cz, ready to serve 200,000 pages this year. It runs customized version of DSpace system with adapted Manakin interface⁸.

The general workflow of the project, shown on *Figure 1* reflects different types of acquired input data:

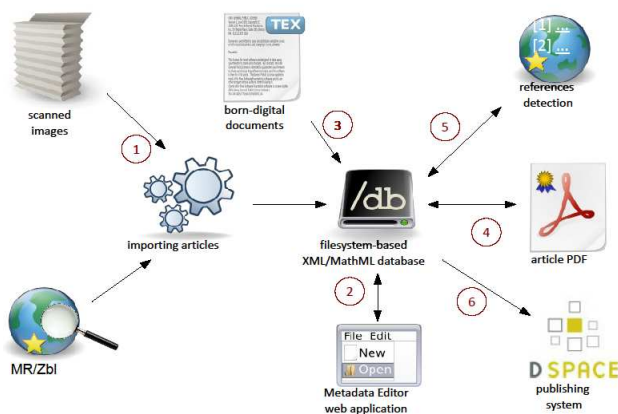


Figure 1 — DML-CZ top-level workflow scheme.

full digitization from prints work starts from a paper copy;

full digitization from bitmap image work starts from an electronic bitmap of pages;

retro-born-digital work starts from an electronic version of the document (usually in POSTSCRIPT or PDF);

born-digital workflow of the journal production is enriched with an automated export of data for the digital library.

⁷ DML-CZ project <<http://project.dml.cz/>>.

⁸ DSpace <<http://www.dspace.org/>>.

DSpace Manakin <<http://di.tamu.edu/projects/xmlui/manakin/>>.

KREJČÍŘ V. Building Czech Digital Mathematics Library upon DSpace Systém.

Within the project, several general purpose tools have been developed, in addition to the DSpace adaptations:

DML-CZ OCR workflow allowing recognition of scanned mathematical documents,

web-based Metadata Editor⁹,

tools for classification of mathematical documents and measuring their similarity;¹⁰

workflow for born-digital publication production with direct export of metadata for DML¹¹ and

plenty of other smaller tools like: extensions to Lucene engine allowing indexing of mathematics, batch PDF stamper for digitally signing of produced PDF, an optimizer recompressing image objects in PDF with the new JBIG compression filter supported by Adobe since PDF specification version 1.6 (Adobe Reader 5) or batch file article PDF generation with titlepage by XeLaTeX.

In the following sections we describe main features of these tools and technologies with the hope that they can be used by similar projects in mathematical or other domains.

DML-CZ OCR

Tests with various OCR programmes showed that no single one gives acceptable results for mathematical content, with character error rates often above 10% (counting wrong character positions and font types as errors too). For text recognition, FineReader by ABBYY¹² gave the best results, whereas for the structural recognition of mathematics InftyReader¹³ had impressive results. We have communicated to the authors of Infty Project the possibility to combine the programmes, and got a version of the programme that is able to read PDF with a text layer inserted by FineReader¹⁴. We found that setting the parameters of the OCR engine (language, word-list consultation) influences the precision significantly. We trained FineReader on the type cases used at the printer where journals were typeset. At the end of extensive experiments, we developed a method of OCR processing consisting of several phases:

1. A page or block of text is recognised for the first time using a universal setup (non-language specific). A histogram of character bigrams and trigrams from words with lengths higher than three is created.

⁹ BARTOŠEK M., KOVÁŘ P. and ŠÁRFY M. DMLCZ Metadata Editor.

¹⁰ ŘEHŮŘEK R. and SOJKA P. Automated Classification and Categorization of Mathematical Knowledge.

¹¹ RŮŽIČKA M. Automated Processing of TEX-typeset Articles for a Digital Library.

¹² ABBYY <<http://www.abbyy.com/>>.

¹³ SUZUKI M. at all. INFTY : An integrated OCR system for mathematical documents.

¹⁴ KANAORI T. and SUZUKI M. Refinement of digitized documents through recognition of mathematical formulae.

2. The computed histogram of the text block is compared to the histograms created from the journal data during the training phase for all languages used (English, French, Russian, German and Czech). Perl module `Lingua::Ident` is used. Block with bibliography is detected by different algorithms and is treated differently.
3. Page or block of text is processed for the second time with parameters optimised for recognised 'language' in previous step and saved as PDF with text layer.
4. PDF is passed to `InftyReader` and results are stored in `Infty Markup Language (IML)`.
5. IML is postprocessed by a home-grown programme in Java to fix recognition errors of some of the accented characters that `Infty` does not yet have in its glyph database.

Using the process outlined above we managed to decrease the character error rate from initial 11.35% (universal language setup of `FineReader`) to an average 0.98% character error rate. The whole processing is fully automated after initial training. Error rate may be decreased further when `Infty`'s character database is semiautomatically enriched when processing a new journal.

Metadata Editor

Metadata Editor (ME)¹⁵ has gradually developed into an efficient web application that allows simultaneous distant editing according to assigned structured access rights. It supports two levels of actions. On the first one the operator editing the data is provided with page thumbnails so that he can visually check the completeness, scan the quality and configuration of the articles, easily shuffle the pages and cut or merge articles if necessary. On the other level the operator can check the automatically imported metadata, edit and complete them. An integral part of the ME is the module for administration of authority files with authors' names. It enables the most suitable version of the name for the DML-CZ to be selected and to match it with all its other versions.

These functionalities in combination with remote access enable to distribute the work among several people on different levels of expertise. GUI allows hired operators (mostly students of mathematics) intuitive work on the entry level. They inspect and correct the structure of complex objects (journal – volumes – issues – articles). Afterwards, they make the initial inspection of the metadata, add the titles in the original languages, provide notes signaling possible problems. Experienced mathematicians then add the necessary translations, complete the missing MSC codes¹⁶, provide links between related papers. They also accomplish the final revision and validation of the metadata.

¹⁵ BARTOŠEK M., KOVÁŘ P. and ŠÁRFY M. *DMLCZ Metadata Editor*.
ME <<http://sourceforge.net/projects/dme/>>.

¹⁶ MSC. *Mathematics Subject Classification* <www.ams.org/msc/>.

We consider bibliographical references as important metadata of every paper. Their availability makes it possible to use professional systems like `CrossRef` for cross-publisher citation linking. The work starts from OCR of the text, in which a block of references is found. Citations are tagged by a script based on regular expressions written for the citation style of every journal. The operator then checks, edits and approves the list of paper citations.

For fixing errors that can be safely detected (as MSC code string invalid in MSC2000) procedures are formulated and coded in XSchema generated also from a developed web-based interface (forms). Other sets of constraint checkers run as overnight jobs together with updates of the database and metadata statistics and logs useful for the management of Metadata Editor workflow.

Finally, various detection procedures of possible errors have been suggested, evaluated and implemented for finding anomalous and suspicious content of metadata fields, with lists of warnings generated including hyperlinks for easy checking by an operator. An important control concerns the integrity of TeX sequences in metadata to assure a seamless typesetting of article cover pages in the later stage: all metadata to be typeset are exported in one big file with unique reference to the article, and typeset by XeLaTeX to check the TeX control sequences used in the metadata fields. This ensures that all of the TeX encoded mathematics converts into the MathML format smoothly. Similar procedures allow for an efficient and economical increase of metadata completeness and quality.

Mathematical Document Classification and Categorization

Fine document classification allows document filtering to reach higher precision in the information retrieval system as DML. The most commonly used classification system today is the Mathematics Subject Classification (MSC) scheme www.ams.org/msc/, Almost all of peer-reviewed mathematics journals use it, but as it has been adopted only in nineties old papers lack these classification tags. We have developed a MSC classifier (guessed MSC) that is able to assign top-level MSC for retro-digitized articles. Our results convincingly demonstrated the feasibility of a machine learning approach to the classification of mathematical papers.¹⁷

Another round of experiments was done with mathematical document similarity computation. We have collected corpus of more than 20,000 journal article fulltexts and we computed paper similarities using *tfidf*¹⁸ and Latent Semantic Analysis (LSA)¹⁹ and Random Projection methods. Methods use a Vector Space Model, first

¹⁷ ŘEHŮŘEK R. and SOJKA P. *Automated Classification and Categorization of Mathematical Knowledge*.

¹⁸ SALTON G. and BUCKLEY Ch. *Term-weighting approaches in automatic text retrieval*.

¹⁹ DEERWESTER S.C. *at all. Indexing by latent semantic analysis*.

converting articles to vectors and then using the cosine of the angle between the two document vectors to assess their content similarity.²⁰ The difference between the methods is that while tfidf works directly over tokens, LSA first extracts concepts, then projects the vectors into this conceptual space where it only computes similarity.

We are now going to show the links to closest document lists in our DML-CZ article pages to get the feedback from authors and readers to evaluate metrics computed in this experiment. It helps to tackle plagiarism, too.

Unifying Metadata

Ways to acquire metadata for articles from different periods (retro-digital, retro-born and born-digital) differ. Some journals have already volume of retro-digital and retroborn periods available in referative databases and import their initial version. For other journals we started from OCR texts and edited them in ME. Metadata editor together with set of transformations (in XSLT) and import filters is indispensable for these types of tasks and their proper timing (ordering) has to be ensured by the software developed.

Most publishers' workflow starts from properly tagged input data (well structured validated LaTeX or MathML). Their workflow could be adjusted only slightly to get proper validated metadata for DML-CZ DL directly as a side-effect of the main publishing process. We have been doing this kind of cooperation with several publishers switching to DML-CZ as their electronic publishing platform: Masaryk University Press for journal *Archivum Mathematicum*²¹, Charles University for *Commentat. Math. Univ. Carol.* (CMUC) and we are working with Academy of Sciences ČR for journals *Math. Bohemica*, *Czech Mathematical Journal*, *Applications of Mathematics* and *Kybernetika* and Palacky University for *Acta Univ. Palacki Olomouc*.

With the developed workflow for the born-digital data files are available in the library almost instantly together with the printed publication, without additional costs.

Another big agenda in the digitization project are issues related to IPR, metadata exports, visibility, search engine optimization (SEO). We have communicated with Google Scholar so that article metadata used in Google Scholar are taken directly (from article's XHTML landing page metadata), and is not guessed from PDF files. This improved quality and precision of article searches considerably. We monitor repository usage via Google Analytics and take care of SEO issues. We also digitally sign exported PDFs do that they are distinguishable from falsehood.

We evaluated costs related to publishing articles in digital form against author's will (publisher mostly have

only written author's consent for printed publication). For scientific papers these are low (primary reason of scientific publishing is spreading the knowledge and it does not matter if in printed form or electronically). No author requested his/her paper to withdraw from repository yet.

Conclusion and Acknowledgement

We believe that the DML-CZ and methods and tools described represent a step towards a European or even world-wide framework for a digital mathematics library, bottom-up evolved from smaller "pilot" projects. This research has been partially supported by the grant reg. no. 1ET200190513 of the Academy of Sciences of the Czech Republic, by MŠMT grants MSM0021622419 and 2C06009.

References

- BARTOŠEK, Miroslav, KOVÁŘ, Petr and ŠÁRFY, Martin. DMLCZ Metadata Editor: Content Creation System for Digital Libraries. In SOJKA, Petr (ed.) *Towards Digital Mathematics Library – Proceedings of DML 2008*, Birmingham, UK, July 2008. Masaryk University, p.139–151.
- BARTOŠEK, Miroslav, LHOTÁK, Martin, RÁKOSNÍK, Jiří, SOJKA, Petr and ŠÁRFY, Martin. DML-CZ: The Objectives and the First Steps. In BORWEIN, Jonathan, ROCHA, Eugénio M. and RODRIGUES, José Francisco, (eds.), *CMDE 2006: Communicating Mathematics in the Digital Era*, A. K. Peters, MA, USA, 2008, p. 69–79.
- BOUCHE Thierry. Next Digital Mathematics Library. In SOJKA, Petr (ed.) *Towards Digital Mathematics Library – Proceedings of DML 2008*, Birmingham, UK, July 2008. Masaryk University, p. 3–15.
- DEERWESTER, Scott C., DUMAIS, Susan T., LANDAUER, Thomas K., FURNAS, George W. and HARSHMAN, Richard A. Indexing by latent semantic analysis. *Journal of the American Society of Information Science*, 41(6) 1990, p. 391–407.
- KANAHORI, Toshihiro and SUZUKI, Masakazu. Refinement of digitized documents through recognition of mathematical formulae. In *Proceedings of the 2nd International Workshop on Document Image Analysis for libraries*, Lyon, France, April 2006, p. 95–104.
- KREJČÍŘ, Vlastimil. Building Czech Digital Mathematics Library upon DSpace System. In SOJKA, Petr (ed.) *Towards Digital Mathematics Library – Proceedings of DML 2008*, Birmingham, UK, July 2008. Masaryk University, p. 117–126.
- MANNING, Christopher D., RAGHAVAN, Prabhakar and SCHÜTZE, Hinrich. *Introduction to Information Retrieval*. Cambridge University Press, 2008. ISBN: 978-0521865715.
- ŘEHŮŘEK, Radim and SOJKA, Petr. Automated Classification and Categorization of Mathematical Knowledge. In AUTEXIER, Serge, CAMPBELL, John, RUBIO, Julio, SORGE, Volker, SUZUKI, Masakazu and WIEDIJK, Freek. (eds.), *Intelligent Computer Mathematics – Proceedings of 7th International Conference on Mathematical Knowledge Management MKM. Lecture Notes in Computer Science LNCS/LNAI*, July 2008, vol. 5144, p. 543–557, Berlin, Heidelberg, Springer-Verlag.
- RŮŽIČKA, Michal. Automated Processing of TeX-typeset Articles for a Digital Library. In SOJKA, Petr (ed.) *Towards Digital Mathematics Library – Proceedings of DML 2008*, Birmingham, UK, July 2008. Masaryk University, p. 167–176.
- SALTON Gerard and BUCKLEY Chris. Term-weighting approaches in automatic text retrieval. *Information Processing and Management*, 24, 1988, p. 513–523.

²⁰ MANNING Ch.D., RAGHAVAN P. and SCHÜTZE H. *Introduction to Information Retrieval*.

²¹ RŮŽIČKA M. *Automated Processing of TEX-typeset Articles for a Digital Library*.

SOJKA, Petr. From Scanned Image to Knowledge Sharing. In TOCHTERMANN, Klaus and MAURER, Hermann (eds.), *Proceedings of I-KNOW'05: Fifth International Conference on Knowledge Management*, Graz, Austria, June 2005. Know-Center in coop. with Graz Uni, Joanneum Research and Springer Pub. Co., p. 664–672.

SOJKA, Petr. *DML 2008 : Towards a Digital Mathematics Library*. Proceedings of DML 2008, Birmingham, UK, July 2008. Brno, Czech Republic : Masaryk University Press, 2009. 149 s. ISBN 978-80-210-4781-5.

SUZUKI, Masakazu, TAMARI, Fumikazu, FUKUDA, Ryoji, UCHIDA, Seiichi and KANAHORI, Toshihiro. INFTY : An integrated OCR system for mathematical documents. In VANOIRBEEK C., ROISIN C. and MUNSON E. (eds.), *Proceedings of ACM Symposium on Document Engineering 2003*, Grenoble, France, 2003. ACM, p. 95–104.

WEGNER, Bernd. RusDML 2008: Current Facilities of the Core Archive of Digitized Russian Publications in Mathematics. In SOJKA, Petr (ed.) *Towards Digital Mathematics Library – Proceedings of DML 2008*, Birmingham, UK, July 2008. Masaryk University, p. 83–86.

Website resources

arXiv [online]. [Cited Jun 7, 2009]. Available from: <<http://arxiv.org/>>.

CEDRAM. Centre de diffusion de revues académiques mathématiques [online]. [Cited Jun 7, 2009]. Available from: <<http://www.cedram.org/>>.

DML-CZ (Czech Digital Mathematics Library) [online]. [Cited Jun 7, 2009]. Available from: <<http://dml.cz/>>.

DSpace [online]. Open source software. [Cited Jun 7, 2009]. Available from: <<http://www.dspace.org/>>.

DSpace Manakin. [online user interface]. [Cited Jun 7, 2009]. Available from: <<http://di.tamu.edu/projects/xmlui/manakin/>>.

ERAM. The Jahrbuch Project. Electronic Research Archive for Mathematics. [online]. [Cited Jun 7, 2009]. Available from: <<http://www.emis.de/projects/JFM/>>.

Euclid. Mathematics and statistics resources online. [Cited Jun 7, 2009]. Available from: <<http://projecteuclid.org/>>.

Mathematical Reviews [online]. [Cited Jun 7, 2009]. Available from: <<http://www.ams.org/mr-database>>.

Math-Net.Ru. All-Russian mathematical portal [online]. [Cited Jun 7, 2009]. Available from: <<http://www.mathnet.ru/>>.

ME. Digitization metadata editor [online]. [Cited Jun 7, 2009]. Available from: <<http://sourceforge.net/projects/dme/>>.

MSC. Mathematics Subject Classification [online]. [Cited Jun 7, 2009]. Available from: <<http://www.ams.org/mathscinet/msc/msc.html>>.

NUMDAM. Recherche et téléchargement d'archives de revues mathématiques numérisées [online]. [Cited Jun 7, 2009]. Available from: <<http://www.numdam.org/>>.

PMC (PubMed Central) [online] : A free archive of life sciences journals. [Cited Jun 7, 2009]. Available from: <<http://www.pubmedcentral.nih.gov/>>.

Zentralblatt MATH [online database]. [Cited Jun 7, 2009]. Available from: <<http://www.zentralblatt-math.org/zmath/>>.

Open Source versus Commercial Solutions for a Long-term Preservation in Digital Repositories

Andrea Fojtů

andrea.fojtu@ruk.cuni.cz

Charles University of Prague. Czech Republic

Abstract: The poster presents substantial characteristics of today's available, well-known open source and commercial solutions for a long-term preservation of digital documents in digital repositories. In the poster the criteria against which the chosen repositories were assessed are shown in a tabular form. The results of the poster survey led to two possible conclusions.

Keywords: open source repository, commercial repository, digital repository, long-term preservation, digital curation OAIS model

A long-term preservation

The common terminology and conceptual framework for all projects dealing with the long-term preservation of digital documents is the OAIS model (the Open Archival Information System). It defines "an archive, consisting of an organization of people and systems that has accepted the responsibility to preserve information and make it available for a designated community [...] and for long enough to be concerned with the impacts of changing technologies, including support for new media and data formats, or with a changing user community".¹

In the Anglo-American information sources we mostly come across terms like digital preservation, (long-term) preservation of digital objects or digital curation.

The importance of the digital preservation may be corroborated by a Rothenberg's famous saying: "the digital information lasts forever or five years, whichever comes first".²

An initial point

An impulse for the survey of open source versus commercial solutions for a long-term preservation was the Repository Software Survey, conducted by JISC (March 2009) and focused on comparison of different aspects³.

In the presented poster the criteria were focused prevalently on long-term preservation: a, existence of OAIS model implementation; b, a wide range of supported formats; c, an open architecture for other applications and plug-ins; d, internal tools for format change (e.g. emulation, migration); e, a SW platform and HW (in)dependence; f, administrators' functions; and g, services.

Comparison & Results

In order it was possible to collate the open source versus commercial solutions (the most often implemented in today's information institutions), SW from the afore-mentioned

study⁴ were chosen: a) open source SW: DSpace, Fedora, EPrints and Research-Output Repository Platform; b) commercial SW: CONTENTdm, Digital Commons, Digitool, Equella, intraLibrary, Open Repository and Vital. Three relatively "new" systems were added to the evaluation, namely IBM Dias, Tessella SDB, Ex Libris Rosetta.

At the beginning two major postulates prevailed: there would be at least one SW solution (open source or commercial) that would be suitable for dealing with the long-term preservation. The second one is related to better performance of open source systems because of a widespread developer and user community.

Yet, as the repositories' solutions were compared none of the abovementioned presumptions were confirmed.

Conclusions

The results of the poster survey lead to two possible conclusions. The first has to do with the very development in the field of repositories for a long-term preservation is still in its infancy and more strengths has to be applied. The latter one has to do with the fact that however the better the system is, human and financial resources as well as risk management (including testing, auditing and certification) play a very important role.

References and Website resources

ISO 14721:2003, Space data and information transfer systems — Open archival information system — Reference model.

Repository Software Survey [online]. JISC RepositoryNet, March 2009 [Cited 2009-04-05]. Available from: <<http://www.rsp.ac.uk/software/surveyresults>>.

ROTHENBERG, Jeff. *Long-term Preservation of Digital Information: Challenges and Possible Technical Solutions* [online]. NEDLIB talk, Dec 2000 [cited 2009-04-05]. Available from: <http://nedlib.kb.nl/workshop/rothenberg_kb.pdf>.

¹ ISO 14721:2003.

² ROTHENBERG J. *Long-term Preservation of Digital Information*.

³ Repository Software Survey <<http://www.rsp.ac.uk/software/surveyresults>>.

⁴ ROTHENBERG J. *Long-term Preservation of Digital Information*.

National Repositories of ETDs and Grey Literature in the Czech Republic

Iva Horová & Jan Mach

iva.horova@amu.cz, mach@vse.cz

Academy of Performing Arts in Prague, Library, Czech Republic / University of Economics, Prague, CILS, Czech Republic

Abstrakt: the poster introduces two Czech large projects that address the question of free access to ETD and gray literature, Electronic Theses and Dissertations Working Group and current issues in this area.

Keywords: repositories; projects; Czech Universities; ETD; metadata

Starting Situation

Until the last year there was no access to Theses and Dissertations (bachelor, diploma, theses and dissertation) on the national level in the Czech Republic. Potential users were left with fragmented sources of varying quality. In 2004 an initiative of university librarians led to the foundation of the Electronic Theses and Dissertations Working Group (<http://www.evskp.cz>) in the framework of the Association of Libraries of Czech Universities (<http://www.akvs.cz/en/>). The ETD Working Group set up goals, made analyses and started systematic work to prepare conditions on local level to establish a national register (guidelines, recommendations, methodology etc, published in web form). The situation significantly changed after 2006 when the Higher Education Act, as amended, provided for an obligation to make ETDs accessible for the public.

Projects

Last year two projects on digital repositories were introduced in the Czech Republic. The **Theses.cz** project (<http://theses.cz/>) was subsidized by the Czech Ministry of Education in 2008 (National Register of Thesis and Plagiarism-Tracing system). This project was managed by Masaryk University Brno in collaboration with 17 Universities of the Czech Republic. Now there are 21 participants. It represents a nationwide registry of theses and also all the information related to these (their names, names of their authors, etc.). The second is **The Digital Library for Grey Literature** (<http://nusl.stk.cz>), project of

The State Technical Library in Prague which is solved in collaboration with University of Economics, Prague. It is planned for period of four years (2008–2011) and it is supported by the Ministry of Culture. The project shall provide a working pilot application.

About ETD Working Group

The Working Group needed to set the metadata used nationwide, allowing publishing in the international repositories. Therefore they prepared three Czech XML standards to describe ETDs including fulltexts, persons and corporations on the basis of DC, DC terms, ETD-MS and XMetadiss formats. The standard prepared as a working paper in 2006, started to be widespread at Czech universities and significantly contributed to harvest the ETDs retrospectively. All the three standards are to be used to describe the Czech grey literature in the Digital Library. The ETD Working Group in cooperation with the Academy of Performing Arts in Prague are working on a solution to integrate the non-text ETDs (audio, video, photo etc.) into the project as an equivalent of text materials.

The poster introduces projects, presents preparation and functions of both repositories and it aims, standards and the metadata used. It also tries to offer the best practices for similar consortium projects, problems and their solutions learned in collaboration on the national level, by preparing licence agreements, workflows and standards used nationwide in the Czech Republic.

Institutional Repository at the University

Iva Prochásková & Lucie Vyčítalová

iva.prochaskova@upce.cz, lucie.vycitalova@upce.cz

University of Pardubice

Abstract: The role of the university library, threats, opportunities and the most important preconditions to set up a repository are mentioned in the introduction. The second part includes a case study at the University of Pardubice.

Keywords: Institutional repositories, scholarly communication, university libraries, DSpace.

Introduction

Not only one definition can be found in the materials published about institutional repositories (IR). The SPARC¹ organization defines IR as follows: institutionally defined, scholarly, cumulative and perpetual, open and interoperable. In accordance with the definition created by C. Lynch² “a university-based institutional repository is **a set of services** that a university offers to the members of its community for the management and dissemination of digital materials created by the institution and its members.” This set of services must be adapted to specific needs of the university. At universities in the Czech Republic these needs can differ from the needs at universities in western European countries and USA.

Who should provide these services? There are fairly good reasons that the university libraries do that. It is a great challenge to enhance their importance for the university especially in our country where university libraries used to be mere lending offices for a long time. There is the possibility to use our knowledge not only for providing the academic audience with the external information but also vice versa: to provide the world with scholar materials created by academic staff and students. Librarians have the best preconditions for these activities. They are familiar with standardized documents processing. They quickly accept new information technologies and, last but not least, they are used to training users in information skills. They are constantly in touch with scholars when providing them with books, articles and other information resources. Therefore, they can easily help them also in presenting their research outputs.

Of course there are some threats as well. A sort of librarians' conservatism could be the first one. Librarians often adhere to the established processes and try to apply new technologies without any serious analysis which could bring a creative change of these processes. The second one can be a lack of good arguments when explaining the

advantages that the functional repository will definitely have for the university. To convince the university management of supporting the setting up a repository requires high level and up-to-date knowledge in such a complex environment that could be called (in a simplified way) the information industry. This knowledge and the motivation are the most important preconditions to a successful departure. Other indispensable conditions are the support of the IT department and some financial resources.

The case of the IR at the University of Pardubice

The main impulsion to consider the IR was the amendment 624/2006 of the Higher Education Act 11/1998³. According to this amendment, the university is obliged to make accessible the submitted and accepted theses and dissertations (TD) to public. The current situation in TD processing and presenting was analysed and the project proposal was prepared. The aim of the project was to set up the IR intended for presentation and archiving the electronic theses and dissertations (ETD) and to integrate it into the university information system (UIS). The project was launched in 2007 with help of a financial support from the Funding Agency for Universities Development. When deciding about the software we took into consideration the experience of the university library at VSB-Technical university of Ostrava where the first Czech installation of the open source software DSpace was put through in 2006 and we have chosen this one as well. This software fulfills our main requests e.g. interoperability, facile installation, large uses and cost-effectiveness. As the first step, we specified the structure of communities, sub-communities and collection in accordance with the needs and structure of our university. Afterwards, the successful installation and Czech localization was carried out. With regard to the Czech metadata standard EVSKP-MS, the templates for metadata entry were prepared. The workflow was created with using the current practice when a student

¹ SPARC <<http://www.arl.org/sparc/>>.

² LYNCH C. *Institutional repositories*.

³ Zákon č. 111/1998 Sb. o vysokých školách a o změně a doplnění dalších zákonů.

(or staff entrusted by a faculty or department) inserts the basic data about the thesis into the UIS for managing all procedures connected with students' studies (STAG). The insertion of ETD in fulltext and the selection of access policy became an integral part of this process. The metadata and fulltexts of defended ETDs are regularly exported from STAG to DSpace.

The second part of the project involved the import of the bibliographic records from the library automated system where the TDs had been processed until then. It needed the preparation of a mapping table between UNIMARC and Dublin Core fields. After this import, the collection of TDs submitted at the University of Pardubice from 1992 is accessible through the IR.

Even though the project was primarily focused to ETDs, the basic conditions for implementing other types of documents into the IR have been created. Since 2009 scientific papers of all faculties are presented there. At the beginning of 2009 the version 1.5.1 of DSpace was installed, the look of the IR has been adapted to the official web presentation of our university and the name Digital Library of the University Pardubice has been chosen as an official name of our IR. The use of the IR is monitored by Google Analytics. The statistical data concerning the way of access show that the indexing of the IR by searching

servers is an essential condition. Only the minority of users access the IR directly from the university webpage. By setting up the OAI-PMH in May 2009, the IR became an open archive and could be registered in the world registers such as OpenDoar or project DRIVER.

The future development is intended towards the presentation of university research outputs. The pilot project will be carried out in 2009. The workflow of the self-archiving will be tested with attendance of two researchers from each faculty with using the current UIS for managing bibliographic data of research outputs (OBD).

References and Website resources

LYNCH, Clifford. Institutional repositories: essential infrastructure for scholarship in the digital age. *ARL : A Bimonthly Report* [online]. February 2003, no. 226 [cited 2009-04-17]. Available from: <<http://www.arl.org/resources/pubs/br/br226/br226ir.shtml>>.

SPARC (Scholarly Publishing and Academic Resources Coalition) [online]. [Cited 2009-04-17]. Available from: <<http://www.arl.org/sparc/>>.

Zákon č. 111/1998 Sb. o vysokých školách a o změně a doplnění dalších zákonů.

Sponsors:



Autorizovaný distributor **3M**



CASLIN 2009
Institutional Online Repositories and Open Access

16th International Seminar
Teplá Monastery, Czech Republic, 7-11 June 2009

Editor Zdenka Firstová
Published in Pilsen 2009

by

University of West Bohemia, Pilsen
Univerzitní 8, 306 14, Pilsen, Czech Republic
(Západočeská univerzita v Plzni, Univerzitní 8, 306 14 Plzeň)
Online available from: <<http://www.knihovna.zcu.cz/Caslin/Caslin09.pdf>>

ISBN 978-80-7043-806-0