Clobridge, Abby

Libraries in transition: from book collections & union catalogues to Open Access & digital repositories

ProInflow. 2011, vol. 3, iss. 2, pp. 121-132

ISSN 1804-2406

Abby Clobridge

LIBRARIES IN TRANSITION:
FROM BOOK COLLECTIONS & UNION CATALOGUES TO OPEN ACCESS & DIGITAL REPOSITORIES

Zajímavosti z oboru

Abstrakt:

Souborné katalogy jsou specifickým typem knihovních katalogů, které shromažďují a poskytují informace z více knihoven prostřednictvím jednoho unifikovaného rozhraní. Souborné katalogy původně představovaly seznamy knih či seriálů knihoven z dané lokality nebo daného oboru. S rozvojem a expanzí digitálního obsahu se objevují souborné katalogy další generace, které shromažďují obsah z různých digitálních repozitářů. Tento článek prezentuje obecný přehled o problematice Open Access, digitálních repozitářů a interoperability a věnuje se budoucnosti souborných katalogů a knihoven obecně.

Klíčová slova: *Open Access, digitální repozitáře, interoperabilita, souborné katalogy, budoucnost knihoven*

Abstract:

Union catalogues are a specific type of library catalogue that aggregate and present information from multiple libraries in a single, unified interface. Union catalogues were originally listings of books or serials collected by a set of libraries in a given geographic area or subject area. As the information landscape continues to evolve with more exclusively-digital content being produced, a next generation of union catalogues is being developed to tie together content from disparate digital repositories. This paper will provide an overview of Open Access, digital repositories, and interoperability, and it will suggest a framework for considering the future of union catalogues and libraries in general.

Keywords: *Open Access, digital repostitories, interoperability, union catalogues, future of libraries*

1 Introduction: Traditional Libraries, Librarianship, and Union Catalogues

Since the Ancient Library of Alexandria was constructed in the 3rd century BC until recent years, libraries have enjoyed a fairly static environment. Supporting the information lifecycle – or, the range of activities surrounding how information is created, disseminated, collected, organized, catalogued, described, and preserved – has always been at the core of librarians' work. Even so, traditional librarianship was mainly focused on printed materials such as books and journals and on tasks related to selecting, organizing, cataloguing, and preserving those objects. But the Internet and developments in technology are completely changing the way we think about information: how we access information; how we use, reuse, manipulate, and work with information and data; who has access to information; where and when information is accessible; and how we can ensure access to digital information in the future.

As we begin grappling with these questions, it is clear that traditional library roles tied to books and buildings are not sufficient in the digital world – a messy, disorganized, uncontrolled space where anyone with Internet access can publish information without intervention by a publisher or the peer-review process. Librarians' skills – specifically, our ability to work with large masses of information, organize it, and present it in meaningful ways – are more important than ever. But we need to quickly shift how we think about our work and start to tackle en masse these new challenges.

Union catalogues were an early and highly-successful method by which libraries took advantage of new technologies to provide value-added services for both users and librarians. Traditional union catalogues are library catalogues that contain information about holdings from different places, all presented through a single interface. For users, union catalogues facilitate access to information by allowing users to search holdings for multiple libraries at once; browse through keywords or subject headings in larger, aggregated masses of holdings; and, at many libraries, see which library has a particular item as a first step in submitting an interlibrary loan request. Within a single institution, union catalogues connect holdings from multiple libraries or campuses. Specialty libraries such as law libraries and local consortia often establish union catalogues as a service to local patrons. National union catalogues present an entire country's holdings in one web site.

Union catalogues are the result of a shared set of values common among libraries: interoperability among systems, interoperability of data through MARC records, and cooperation among participating libraries. They also require a shared goal of

facilitating access to information for our users and doing what we can to create a seamless environment in which users can access information regardless of its physical location.

With the transition to a digital environment, this need to aggregate and present information from a variety of places within a single interface still exists – although users' expectations from such services are more complex. Traditional union catalogues were the first step in getting access to a physical item; with "next generation" union catalogues, users are able to access items themselves. The information ecosystem has changed – and it is going to continue changing at an exponential rate. Libraries' first step in working in this new environment has been to create a network of distributed repositories and associated services designed to collect, archive, disseminate, curate, describe, and preserve digital materials. These Open Access repositories along with the tools to aggregate their content are the next step in the evolution of union catalogues; it is an important step in the development of libraries as we move into the digital landscape.

2 Open Access & Digital Repositories: An Overview

"Open Access" refers to the practice of granting free access to scholarly research via the Internet. One of the most comment definitions notes that "Open-access (OA) literature is digital, online, free of charge, and free of most copyright and licensing restrictions. What makes it possible is the internet and the consent of the author or copyright-holder."[1]

Open Access can occur via two methods:

1) "Gold" Open Access: Achieved by publishing in any Open Access journal, a peer-reviewed, scholarly journal in which articles are freely available online.

2) "Green" Open Access: Achieved by publishing in any peer-reviewed journal and then depositing a peer-reviewed version of the article in an Open Access repository. Open Access repositories can include articles that were published in gold Open Access journals or closed-access journals.

The peer-review process is critical for either method.

Furthermore, Open Access includes two level of "openness," *gratis* versus *libre*. *Gratis* Open Access provides access to content at no cost other than the costs associated with accessing the Internet. The copyright holder retains all permissions

1 SUBER, Peter, "A Very Brief Introduction to Open Access", at: http://www.earlham.edu/~peters/fos/brief.htm [accessed 30/05/2011].

except for the right to self-archive (i.e., deposit into a personal or institutional repository). On the other hand, *libre* Open Access includes a much wider range of rights and permissions for how others can use the content. In addition to granting permission to access an object, *libre* Open Access allows others to remix or edit content, depending on the exact details of the license associated with a particular item.[2] *Libre* Open Access is often indicated with specific types of Creative Commons licenses such as the "Attribution" or "CC-BY" license in which the copyright owner allows others to "distribute, remix, tweak, and build upon your work, even commercially, as long as they credit you for the original creation. This is the most accommodating of licenses offered."[3]

The motivations for supporting Open Access are diverse but all have increased access at their core. Some common themes:

- **Increase access to scholarship**. By making current scholarship freely and quickly available, it will reduce the access barriers for researchers working outside of well-funded higher education institutions in the developed world. Most researchers working in developing countries and countries in transition – along with researchers in the developed world who work at small and medium-sized enterprises – have minimal access to current scientific journals. Open Access provides one way to level the playing field and provide free access to current scholarship to anyone with an Internet connection. Faster and increased access to research has the potential to spark innovation, creativity, and the production of new throughout the world.
- **Use Information and Communication Technology (ICT) to further disseminate research and allow scientists to conduct research on scholarship.** By opening repositories to search engines, the potential for discovery is far greater than if articles are only available through proprietary publisher-owned databases. The search behaviour of students and researchers has dramatically changed over the past ten years; users tend to go to a search engine first before turning to publishers' databases – even if they have access to publisher databases. In terms of research and development, new types of data analysis make it possible for scientists to study research outputs in ways that are impossible in closed networks or with analogue data. Plus, being able to study data from large aggregates of information potentially can be more valuable than running the same tests on smaller sets.
- **Provide public access to publicly-funded research.** A great deal of research is financed by grants from publicly-supported research funding

2 SUBER, Peter, "An Introduction to Open Access", at http://www.earlham.edu/~peters/fos/overview.htm [accessed 30/05/2011].

3 "About the [Creative Commons] Licenses", at http://creativecommons.org/licenses [accessed 16/05/2011].

agencies. In the traditional publishing model, research published in journals is only accessible through costly journal subscriptions or electronic databases licensed by libraries. However, with rising costs of electronic resources, many libraries are being forced to cancel subscriptions and limit the journals to which they provide access. No single library anywhere in the world provides access to all journals. Even under the best of circumstances when the public has access to university libraries, many libraries are being forced to restrict public users from accessing electronic journals due to rising subscription costs. Several major research funding organizations have taken this idea one step further and have implemented policies requiring recipients of their grant funds to deposit research resulting from their funding into Open Access repositories. Research funding bodies mandating Open Access include organizations such as the European Commission, the National Institutes of Health (United States of America), and the Wellcome Trust (United Kingdom). Many other public and private research funding agencies have policies in place or under development. Furthermore, several countries including Denmark and Spain are debating implementing national-level policies.

- **Enhance visibility**. Open Access has the potential to enhance the visibility of the research outputs of individual authors, institutions, countries, and regions. Most repository systems come with standard statistics packages so authors can see information about article downloads. Eventually, authors will be able to trace citations of their research in other publications. Through one interface, institutions are able to collect and disseminate all peer-reviewed articles as they are published, making it easier to showcase research and assess the research output of individuals, departments, and the institution as a whole. At the national and regional level, countries are beginning to aggregate content in various configurations such as all of the scholarly output of a given country or a funding agency.

The Open Access movement started in the early 2000s with the Budapest Open Access Initiative (2001), a meeting which led to use of the phrase 'Open Access,' its formal definition, and one of the original declarations in support of Open Access, now an open document that can be signed by the public. In 2003, the other two formative meetings were held in Berlin, Germany and Bethesda, Maryland in the United States that led to further public statements in support of Open Access. Currently, institutions around the world maintain nearly 2,000 repositories[4], and over 6,500 Open Access journals (including over 570,000 articles) are produced by publishers

4 "The Directory of Open Access Repositories", at http://www.opendoar.org [accessed 30/05/2011].
"The Registry of Open Access Repositories", at http://roar.eprints.org [accessed 30/05/2011].

and other organizations[5]. Open Access has been most widely adopted within Europe and North America, although increasing activity is being seen in Latin America, Asia, and Africa. For developing countries, Open Access is appealing in two ways: to increase access to current research that was previously inaccessible and to disseminate locally-produced research on a global scale.

Even in the ten years since the launch of Open Access, the information ecosystem has changed. While the initial focus was on peer-reviewed scholarly articles (postprints), digital repositories can handle a wide variety of content types, and the boundaries between Open Access and other forms of scholarly materials held in digital repositories is blurring. Currently, many repository managers are examining the feasibility of incorporating enhanced publications (published research plus the associated data sets, documents, models, images, etc. associated with particular publications) into repositories. A wide variety of information and content exists in many repositories: grey literature, including conference proceedings and pre-prints; student scholarship such as electronic theses & dissertations (ETDs); and multimedia files – audio, video, digital images.

Each institution needs to be able to collect, curate, disseminate, and preserve the intellectual capital created at that institution, and so each institutional repository has value on its own. But much in the way that traditional union catalogues provided significant value to users by presenting holdings from multiple libraries in a single interface, the real value of Open Access and digital repositories lies in the potential to aggregate research outputs, present information in different ways, and allow for new types of data extraction, data mining, visualizations, and other forms of analysis.

In this regard, some of the earliest and largest repository harvesting projects both clearly identify with the legacy of union catalogues – they function in much the same as traditional union catalogues, but they tie together mainly full-text content from digital repositories, thus providing direct access to digital objects. These Open Access repository harvesting projects are truly a next generation of union catalogs, building on the same principles but going a step further.

One example, OAIster is described as "a union catalog of millions of records representing open access resources that was built from open access collections worldwide using the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). Today, OAIster includes more than 25 million records representing digital resources from more than 1,100 contributors."[6] The OAIster project was developed in the early 2000s at the University of Michigan; in 2009, it was transitioned to OCLC, who still maintains the service. Another significant repository harvesting project, the Networked Digital Library of Theses and Dissertations (NDLTD), also describes its

5 "The Directory of Open Access Journals", at http://www.doaj.org [accessed 16/05/2011].

6 "The OAIster Database", at http://www.oclc.org/oaister [accessed 30/05/2011].

harvesting repository as a union catalog: "The NDLTD Union Catalog contains more than one million records of electronic theses and dissertations. For students and researchers, the Union Catalog makes individual collections of NDLTD member institutions and consortia appear as one seamless digital library of ETDs."[7]

3 Interoperability

OAIster, the NDLTD Union Catalog, and all other tools and services that tie together content from multiple repositories are possible because of interoperability, or the ability of systems to pass information back and forth between them in a usable format. Interoperability is the mechanism by which repositories and other systems, including traditional library catalogues, are able to work together.

To facilitate repository interoperability, most Open Access repositories use Dublin Core, an international standard, as the basis for the structure of their metadata. Dublin Core is the most generic, most flexible, and least granular of the commonly-used metadata standards; as a result of its generalness, it is used for a variety of purposes including not only repositories but also metadata tags coded into the HTML of many web pages.

Dublin Core comes in two varieties. Unqualified Dublin Core includes fifteen elements or fields, covering basic descriptive and administrative information: "Title," "Creator," "Subject," "Description," "Publisher," "Contributor," "Date," "Type," "Format," "Identifier," "Source," "Language," "Relation," "Coverage," and "Rights." The fields are vague and, in many cases, lack consistency in how they are used from one implementation to another. In order to help alleviate some of the confusion often inherent in Dublin Core, additional "qualifier" elements were added to refine or give more specificity to fields. For example, qualifiers for the "Date" field include: "Created," "Valid," "Available," "Issued," and "Modified." While the qualifiers are often useful to differentiate important bits of information, interoperability applications are tied to unqualified Dublin Core – so enough information needs to be conveyed in the standard 15 elements that an object is findable and usable.

OAI-PMH works by having metadata from a cooperating group of repositories exposed to a harvesting system. Harvesting systems extract the data, either through ongoing, updated processes or through a one-time ingest. Users search the harvested collection of metadata but then access the digital objects themselves from their original home repositories. Only the metadata is extracted from participating repositories, not the digital objects.

OAI-PMH was developed to serve as a low-barrier mechanism for interoperability – and indeed, now nearly all Open Access repositories include OAI-PMH functionality.

7 "Find ETDs – NDLTD", at http://www.ndltd.org/find [accessed 30/05/2011].

Because of its simplicity, OAI-PMH has served as the basis for some of the most widely-adopted interoperability guidelines such as OAIster, the NDLTD Union Catalog, the Digital Repository Infrastructure Vision for European Research (DRIVER) Project, and the Open Access Infrastructure for Research in Europe (OpenAIRE). In all of these cases, a new interface is built as a layer on top of several repositories, bring together content from different places – very much following the model of union catalogues.

Interoperability can also help bind together objects that are somehow related to each other. In the past, scientists conducted research, collected data, and then wrote and published articles. Articles might have included some graphs, charts, drawings, or other affiliated images in limited quantities. Once published, that article became a static entity, and there was not a way to package other related items with a particular article. The digital environment has changed this mindset; researchers are now able to deposit an article in a repository alongside other related items – data sets, algorithms, scripts, transcribed or translated documents, photographs, drawings, charts, graphs, etc. We can even go one step further and create new materials related with research – video or audio interviews with authors discussing their publications, videos of conference presentations, slides from presentations related to the published research.

Curating these groups of objects therefore requires new tools. An interoperability project currently under development is trying to tackle this scenario. The Open Archives Initiative Object Reuse Exchange (OAI-ORE) project is designed to "define standards for the description of aggregations of Web resources." The emphasis of OAI-ORE is on this type of information – these "enhanced publications" or compound digital objects, i.e. the entire body of materials tied to a specific publication such as an article plus its data set; related images, audio, or video files; charts, graphs or visualizations; code for software; etc. With OAI-ORE, the entire aggregate of objects and their associated metadata can be passed back and forth between systems as a unit.

Interoperability among repositories can be used to develop systems and tools that offer more functionality than traditional union catalogues. Another current project, Simple Web-Service Offering Repository Deposit (SWORD), is designed to support authors who are trying to contribute content to repositories. With the introduction of repositories, the publication process has been extended – and become more complicated. Now, after going through the peer review and editing process for journal publication, authors are being asked (or required) to deposit copies of articles into their institutional repositories. If an institutional Open Access policy is in place, authors might need to either attach a waiver or amendment to their copyright agreements at the point when they sign over copyright. They might need to identify

which version of an article meets the publisher's criteria for deposit into an institutional repository. Then they need to deposit the article in one or many repositories, depending on grant funding criteria. Deposits usually require a few steps plus creating metadata and possibly applying taxonomy terms. SWORD is one mechanism to try to facilitate this process, making it as simple and seamless as possible, by creating a single interface and having articles and their associated metadata deposited into multiple, pre-selected repositories.[8]

At the system level, interoperability is necessary for different repositories or other types of systems to be able to pass data or digital objects from one system to another. Each institution is responsible for its own repository, creating a distributed environment. Most institutions use one of a handful of types of systems, most of which are open source, for their repositories: DSpace, ePrints and Fedora are three commonly-adopted systems; Greenstone, CDS Invenio, and bePress Digital Commons are also all fairly common. Many Open Access journals use Open Journal System (OJS). Of all of these systems, Digital Commons is the only one that is not open source. Even so, all of the systems handle typical processes and workflows in strikingly different ways, making it challenging to create service layers that work consistently for all systems.

In addition, many institutions have multiple, inter-related systems that serve different needs: Open Access or institutional repositories usually maintained by libraries; courseware, possibly in conjunction with a repository of learning objects or Open Educational Resources (OERs) maintained by academic technology departments; and, increasingly, Current Research Information Systems (CRIS) such as euroCRIS, that provide administrative frameworks for managing information about awarded grants, funding, and project information as well as project data. Within a single institution, it is likely that several of these systems need to be able to share information with each other, or that one day a union catalogue might be created to look at different types of digital objects from multiple types of systems.

While challenging, in order to attain our goals with Open Access, it is necessary for these systems to be able to share information and pass objects as well as metadata back and forth. The digital environment should allow for us to do things that we couldn't do before. We're just starting to understand what this means. OAI-PMH, OAI-ORE, and SWORD are just three examples of ways that interoperability allows us to connect repositories to each other. The purposes of the three protocols are vastly different, but ultimately they are services or tools that make it possible for our users to interact with digital objects in unprecedented ways – either at the point of discovery for end users (OAI-PMH or OAI-ORE) or earlier in the information lifecycle as content owners are disseminating their objects (SWORD). If our role as

8 "SWORD", at http://swordapp.org [accessed on 17/05/2011].

librarians is to support the information lifecycle, interoperability allows us to develop tools to facilitate and further these processes in ways not even imaginable a few years ago.

4 Moving Forward

Librarians have long understood the importance of our role in creating tools to facilitate access and discovery of information resources, and union catalogues in various formats and guises have been one way to tie together resources from disparate collections. If information exists, but users aren't able to find it, what good it is? Within the proliferation of digital content, this same principle holds true. If libraries are advocating for Open Access and investing significant resources into building, curating, and sustaining digital repositories, then we also need to develop the tools that enhance how we are able to discover objects, access them, and ultimately, how we are able to use, re-use, or adapt them for new purposes.

Supporting Open Access and other digital collections is the next stage of library work, a stage that is directly tied in with the traditions established by union catalogues. But while union catalogues were traditionally supported by a few librarians or small library systems departments, it is time to reconsider all of this work in light of much larger changes within libraries. Technology, user expectations, and the global information ecosystem continue to change at breakneck speeds. While we can't predict the future, we can look at what we've learned from the traditions of union catalogues and the current landscape and apply these lessons to libraries and information science in general so that we can be better positioned for the future.

The key is to let go of the traditional images of libraries as books and buildings and the traditional organizational structure of staff of libraries (public services and technical services). Instead, we need to take a step back and consider bigger-picture questions – and then let that drive strategic planning, resource allocation, organizational development, and decision making. The core of our work is not about books and buildings, the two images most closely associated with libraries, but rather it is about facilitating all aspects of the information lifecycle: creating new knowledge; curating or collecting, organizing, cataloguing and describing information; disseminating information; connecting people to information; and preserving information.

A few points to consider:

1. The information landscape is global. Our users are no longer exclusively our local constituencies; we have a global audience of scientists, researchers, scholars, students, and lifelong learners who comprise our user base. Serving a global community creates new challenges and opportunities – providing full-service remote access and 24/7 access; understanding the realities and limitations of on-the-fly translation services in a search and retrieval context; working with national-level intellectual property rights (IPR) and copyright legislation; working with a wide range of technology hardware, software, and high-speed access to the Internet.
2. We need to move past artificial boundaries. Within the library world, we have created a number of boundaries that are either arbitrary or invisible to outside users – boundaries that complicate how we present access to information or services. For example, we need to work to move past silos separating collections; between systems; between academic disciplines; between libraries, museums, and archives; and between departments – both within libraries (public services vs. technical services) and outside of libraries (libraries vs. information technology). In most instances, these boundaries are either arbitrary or invisible.
3. We are ***all*** information consumers, producers, and collectors. Knowledge creation is no longer predominantly the domain of formally-credentialed scientists and researchers. Rather, people all over the world of all ages are working with information in ways that were unprecedented even ten years ago. The environment will change again in the next ten years. What we do now should be to prepare for the future. We need to build systems, services, and infrastructures that are nimble, agile, modular, standards-based, and interoperable.
4. We need to facilitate and prioritize discoverability and usability of content, not simply access. In the digital world, libraries have begun the arduous process of collecting materials, but the importance of good metadata must not be overlooked. Good metadata – well-described tags, exposed data, well-disseminated data – is the difference between being able to find a relevant object and having it sit unused in a repository. Likewise, once materials are found, how can they be used? What are the technical restrictions? Legal restrictions? How can we better advocate for open content and allowing users to adapt, repurpose, or remix content? Is this a role librarians should be embracing? How can we develop systems that will allow users to interact with information in new ways? What information

problems are keeping our users awake at night – and what can we do to help solve those problems?

It is important that we keep at the forefront of our minds the fundamental purpose of our work and reflect on what we've learned as the world around us changes.

5 Conclusion

Libraries are in the midst of a dramatic period of change. We need to rethink our roles, embrace the changing information ecosystem, separate our professional identities from books and buildings, and focus on new ways to work with information in this changing environment. We need to remain consistent with the heritage and values of libraries and information science from the past 2000 years but embrace new technologies and think more broadly. The history of union catalogues is a useful blueprint and incorporates key values of cooperation, openness, and interoperability, but it is time to think more broadly and boldly and embrace the new information management questions being raised by the continually-evolving digital landscape.

References

1. "About the [Creative Commons] Licenses", at http://creativecommons.org/licenses [accessed 16/05/2011].
2. "The Directory of Open Access Journals", at http://www.doaj.org [accessed 31/05/2011].
3. "The Directory of Open Access Repositories", at http://www.opendoar.org [accessed 31/05/2011].
4. "Find ETDs – NDLTD", at http://www.ndltd.org [accessed 30/05/2011].
5. "The OAIster Database", at http://www.oclc.org/oaister [accessed 30/05/2011].
6. "The Registry of Open Access Repositories", at http://roar.eprints.org [accessed 30/05/2011].
7. SUBER, Peter, "An Introduction to Open Access", at http://www.earlham.edu/~peters/fos/overview.htm [accessed 30/05/2011].
8. SUBER, Peter, "A Very Brief Introduction to Open Access", at http://www.earlham.edu/~peters/brief.htm [accessed 30/05/2011].