Metadata Value Chain for Open Access Journals

Presenter: Holly Mercer (Texas A&M University)

Recorder: Sharon Dyas-Correia (University of Toronto)

In this session, Holly Mercer presented a case for finding global solutions to improve the metadata that is available for journals, particularly small, independent open access journals. She discussed the scholarly communication lifecycle and described how in her view the metadata value chain underpins the scholarly communication system. Examples of the importance of metadata to discovery, access, and use of publications were provided. Suggestions of how librarians can help editors and publishers enhance discoverability and usability for patrons were provided.

INTRODUCTION

Holly Mercer began her presentation by indicating that her focus would not be exclusively on open access journals but would include independent, single-title society journals and journals that receive financial support from an academic department or center. Some titles are 100 percent open access and began that way, and others may be transitioning from print to digital and possibly becoming open access. The common denominator is a scholarly product. Mercer indicated that her presentation would include a brief outline of her background, a presentation of the scholarly communication lifecycle, discussion of the value of metadata, some examples from the presenters experience working with journals and journal editors, and then some conclusions.

BACKGROUND

The speaker pointed out that she does not have an extensive serials or publishing background. She has some background working with metadata, experience doing lots of other things like digital library projects and she has worked mostly with faculty and research staff. Mercer is currently head of digital services and scholarly communication at Texas A&M University Libraries. She is responsible for the efforts of a small team that collaborates with digital initiatives (library information technology) staff to: promote the effective use of scholarly services and the repository; develop repository policies, procedures, workflows, and metadata standards; enhance awareness of the changing landscape of scholarly communication and

intellectual property rights; and evaluate and maintain quality control of the processes,

collections and services provided by the team. She is responsible for supporting faculty in the

use of Texas Digital Library (TDL) tools and services, including scholarly blogs, research wikis,

journal publishing and online conference management.[1] Prior to joining Texas A&M, Mercer

held several positions at the University of Kansas. All of the presenter's experience with

institutional repositories and electronic journal publishing has been with DSpace and Open

Journal Systems (OJS), respectively. She indicated that there are certainly other options, but her

examples would involve DSpace and OJS because they were used both at Texas A&M and the

University of Kansas.

Mercer works a lot with faculty and researchers who want to increase access to

scholarship. She indicated that faculty often confess that they do all their research online

(rarely in the library) in Google or Google Scholar, and they say that they can find most of what

they need full-text and online. The presenter suspects the faculty members really begin their

research with Google, and then move to scholarly resources licensed by the libraries and she

admits this too is sometimes her own practice. Mercer pointed out that a print only journal

may lose out in this scenario. If print has no online presence it will not be discovered. Also, she

indicated that some smaller journal editors and publishers may have a sense of being left out.

Mercer works with metadata at the "item level," and that means article level. It is

almost always some variation of Dublin Core, and the quality can vary greatly. But now article

information from subscription-only journals is available from search engines, and open access

journal content is indexed in subscription databases. Rather than talk about article-level

metadata she revealed the focus of the examples she would present were on holdings and

availability information and how it's presented to library users. She indicated that this issue is important not just to library users, but to journal editors who want discovery and use to be easy.

SCHOLARLY COMMUNICATION LIFECYCLE

The presenter chose to frame the discussion by presenting Bo-Christer Björk's Scientific Communication Life-cycle Model which is a model of the scholarly publishing system including the research and dissemination process. The model focuses on activities, inputs, outputs, controls, and mechanisms.[2] The ScienceModel.net website provides the following description: "The scope of the model is the whole scientific communication value chain, from initial research to the assimilation of research results to improve every-day life. The model treats both informal and formal communication, as well as the publishing of data, but the major focus is on modelling the publishing and indexing of traditional peer reviewed journal articles, as well as the activities of readers to find out about them and access them. The new business models and parallel functions enabled by the Internet, such as open access journals and e-print repositories, are also in focus."[3] Björk presents his model as a model for scientific communication but Mercer suggests that the model is applicable to all scholarly communication.

At the highest level of the model, the boxes, or nodes, equal activities in scholarly communication. At this level there are four major activities including funding research and development, performing the research, communicating the results and applying the knowledge

gained. Mercer chose to focus her presentation on activity number three, communicating the results. The communication can be informal or formal through activities like publication, and can include data-sharing which is currently gaining a lot of attention. In addition to a node for communicating the results, Björk's model also includes a further node for facilitating dissemination and retrieval of results both globally and locally, and for preservation functions. The presenter suggested that these are library functions, inputs and outputs. Metadata standards facilitate information retrieval globally and libraries purchase access to publications for their users (or they are available open access). Libraries also preserve publications, or contract with others to do so.

Looking more closely at global information retrieval node, one sees that in order to facilitate retrieval globally, publications are bundled and distributed into e-services by "infomediaries" or made openly available on the web. The speaker pointed out that metadata is necessary even for full-text searching and it is important for open access as well. Mercer drew attention to the fact that in further nodes, the model includes boxes to demonstrate that metadata are integrated into various search services. Librarians therefore facilitate at the global level as well and metadata are different in each scenario. Indexes in edited bibliographic indexes, indexes in web harvesters for scientific content and indexes in general web search indexes are all included. The least controlled is a general search engine index and the most controlled is an edited bibliographic index.

VALUE OF METADATA

Mercer stated that according to Wikipedia, "A value chain is a chain of activities for a firm operating in a specific industry. Products pass through all activities of the chain in order, and at each activity the product gains some value. The chain of activities gives the products more added value than the sum of values of all activities."[4] The scope of Björk's model is the whole scientific communication value chain, from initial research to the assimilation of research results to improve every-day life. Mercer indicated that the metadata value chain is part of the larger system. Although researchers might perceive they do all their research using internet search engines and that they rely only on full text searching, metadata are critical to discovery and proper citation. In Mercer's view the metadata value chain underpins the scholarly communication system.

CASE STUDIES

Mercer presented several case studies from her work at the University of Kansas and Texas A&M to demonstrate some of the issues that arise for journal editors and publishers. The cases provide examples of how librarians can assist those concerned with providing better access to scholarship. The first case was of an unnamed mathematics journal. The first inquiry Mercer received regarding the journal was in October 2009 and she also attended a meeting in February 2010. The journal is still in the planning stage but there have been many questions concerning open source, creative commons licenses, marketing, promotion, ISSN's and E-journal and OA journal lists were all important to the faculty involved. The Library has been involved in the development of this journal from very early in the process.

The next example given was of the presenters involvement with a small social studies journal called *American Studies with American Studies International.*[5] The editors were sent contracts by database vendors and were uncertain whether they should enter into any contracts. They had been planning to make journal issues available open access with a five year rolling embargo. The editors wondered if database vendors would agree and the answer was in the contract. The editors also wanted to use the submission and review workflows in the Open Journal System (OJS) and they wondered what their competitor journal was doing. They really wanted full text available from Project Muse and thought about digitized, open access to back issues as a promotional tool for the journal. Basically, they thought a lot about what to do and had tried several options. As a result, from 1971 to 2004 the journal was available open access from the journal website, from 2004 to 2007 it was available in a major humanities index and from 2007 onwards it is available in Project Muse. For this title, as in the case for many others, there is no clear comprehensive listing of electronic options in the catalogue for the patron. Open access to portions in lists can be confusing or absent. Some libraries, including Texas A&M, only include open access titles in their A to Z lists when it is requested.

The third example given was for the title *Biodiversity Informatics.*[6] It was already published open access using the OJS and established in the Directory of Open Access Journals (DOAJ). The editor understood metadata and had thorough, good metadata, but the OJS instance was on a server maintained by the editor, and did not benefit from regular backups, updates, and the like. University of Kansas Libraries were looking for opportunities to support campus journals and asked the editor whether he would be willing to transfer to the libraries' new installation of OJS. It is a good fit because the libraries handle the server and system

support and upgrades and the editor focuses on running the journal. It is simple because there is no print version, there have been no title changes and the journal has been listed in the DOAJ for several years.

Seventeenth-Century News was offered as an example of a journal that is online and open access but earlier volumes were in print by subscription.[7] There was a transition period from 2002 to 2006 when the journal was published in both print and online formats. The journal workflow features of OJS were not required because only commissioned reviews were published. Metadata creation and submissions were handled by the journal staff. There were links for interlibrary loan and for retrieve from shelf services. A lot of options were available in the catalogue for retrieval of the electronic version. Five different links with varying holdings appeared in the catalogue and they are confusing to the patron.

The example presented as case number five is of Social Thought and Research, a student-run journal has been published since 1965.[8] It has had several title changes including KS Journal of Sociology, Mid-American Review of Sociology, and Social Thought and Research. Users can subscribe to the print version or the use the digital open access version. Full text is also available via a major vendor and it was formerly indexed in a major abstracting publication. Article metadata is created by journal staff and reviewed by libraries before it is made publicly available. The freely available issues are not linked in Texas A&M University Libraries but are correct at University of Kansas Libraries site. However, the electronic versions of previous titles are not reflected anywhere. The presenter pointed out that all libraries do not necessarily do things the same way and librarians have not always been as thoughtful about metadata as we should be.

Texas Water Journal was the final example provided.[9]   The Texas Water Journal is new

work and one of the editors is assistant director of the Texas Water Resources Institute (TWRI).

The TWRI already has several collections in the Texas A&M Digital Repository, including a

newsletter. They are familiar with TDL services and have been working with the libraries for

several years. TWRI is also involved in statewide efforts to create a Texas Water Digital Library,

hosted by the TDL. There are opportunities to link data sets from TWDL to articles published in

the journal, so description of journal articles and underlying data are important. The journal is

still a work in progress and not quite ready for prime time.

CONCLUSIONS

Mercer indicated that from her examples it is evident that librarians can help in several

ways. There are many roles in hosting or distribution. Libraries can choose to act as

distributors or publishers via an institutional repository or journal publishing platform.

Librarians can explain what contracts mean but must be careful not to assume an advising role

related to contracts. They can assist journal staff in understanding the implications of title

changes and make suggestions on how to treat format changes like the transition from print to

electronic. Librarians can also direct editors and publishers in how to acquire ISSNs or Digital

Object Identifiers (DOIs) from CrossRef. They can also advise on how to get in e-journal lists,

provide OCLC and library catalogue information, and offer assistance related to modifying and

enhancing MARC records. Librarians can explain funder or institutional open access policies

and implications for journals and help editors and authors understand policy.

The presenter asked, "But what about metadata?" She indicated that there are obviously many shared concerns related to discovery, access and use and metadata value is additive. Metadata change across the scholarly publishing value network is important. We must collectively problem solve at the global, not local level and work to improve the metadata that is available. Linked data globally may provide some solutions and librarians can help editors of small, independent open access journals. More conversations between serials librarians, electronic librarians and all the other stakeholders are essential. On May 21, 2010 the World Wide Web Consortium (W3C) announced the creation of the Library Linked Data Incubator Group which may provide some solutions to issues of global interoperability of library data.

QUESTIONS

The session ended with several thoughtful questions from the audience including questions concerning how faculty members perceive the presenters role and how the presenter felt there services compared to those offered by commercial publishers. The response was that the services provided were well received by some faculty and that the team are not publishers but offer some services. When asked what advice she could give on promoting services, Mercer responded that similar to promoting other services, providers need to identify stakeholders and outreach to them.

NOTES

[1] The Texas Digital Library is a 19 member organization of higher education institutions in *Texas.*

[2] *Bo-Christer* Björk. *Scientific Communication Life-cycle Model.* (Helsinki, Finland: Swedish School of Economics and Business Administration) http://www.ScienceModel.net/ (accessed July 16, 2010). Description and series of reusable slides licensed under the Creative Commons Attribution-NonCommercial 1.0 License. http://creative commons.org/licenses/by-nc/1.0/fi/

[3] Bo-Christer Björk. *Scientific Communication Life-Cycle Model.* (Helsinki, Finland: Swedish School of Economics and Business Administration) http://www.ScienceModel.net/ (accessed July 16, 2010).

[4] "Value Chain." *Wikipedia: The Free Encyclopedia*. (San Francisco, CA: Wikimedia Foundation, Inc.) http://en.wikipedia/wiki/Value_chain. (accessed July 16, 2010). Licensed under Creative Commons Attribution-ShareAlike License.

[5] *American Studies with American Studies International*. (Lawrence, Kansas: Midcontinent American Studies Association, University of Kansas, and the Hall Centre for the Humanities). ISSN: 0026-3079.

[6] *Biodiversity Informatics.* (Lawrence, Kansas: Biodiversity Research Center, University of Kansas). ISSN: 1546-9735.

[7] *Seventeenth-Century News.* (College Station, TX: English Department, Texas A&M University). ISSN: 0037-3028.

[8] Soc*ial Thought and Research.* (Lawrence, Kansas: Sociology Department University of Kansas). ISSN: 094-5830.

Holly Mercer is Head of Digital Services and Scholarly Communication at Texas A&M University Libraries, College Station, Texas.

Sharon Dyas-Correia is Head of Serials and Collections Librarian for Open Access Initiatives at University of Toronto Libraries, Toronto, Ontario.