

June 2016

Open Access Metadata for Journals in Directory of Open Access Journals: Who, How, and What Scheme?

Lisa E. Cheby

San Jose State University, lec2937@gmail.com

Follow this and additional works at: <https://scholarworks.sjsu.edu/slissrj>

 Part of the [Cataloging and Metadata Commons](#), [Scholarly Communication Commons](#), and the [Scholarly Publishing Commons](#)

Recommended Citation

Cheby, L. E. (2016). Open access metadata for journals in Directory of Open Access Journals: Who, how, and what scheme? *SLIS Student Research Journal*, 6(1). Retrieved from <http://scholarworks.sjsu.edu/slissrj/vol6/iss1/4>

This article is brought to you by the open access Journals at SJSU ScholarWorks. It has been accepted for inclusion in School of Information Student Research Journal by an authorized administrator of SJSU ScholarWorks. For more information, please contact scholarworks@sjsu.edu.

Introduction

Shrinking library budgets, rising vendor profits, ethics of freedom of information in scholarly research, and advancements in digital archiving and information retrieval have spurred the open access (OA) movement. While OA most often implies an open system of publishing and licensing, there are really two components to OA: publication and metadata. OA publications seek to create outlets for publishing that are free to access outside of databases or other costly modes of distribution. OA metadata seeks to make data about any type of resource in all types of institutions available for harvesting. While OA metadata is necessary to make OA publications discoverable, OA metadata may be created and released for content that is not published as OA. Various schema and crosswalks have been developed in efforts to make metadata interoperable. Additionally, in order to facilitate OA distribution, initiatives to define procedures for creating OA content have emerged from institutions including Harvard and The National Information Standards Organization (NISO).

While some schemes and protocols seem to be emerging into possible standards, there seems to remain confusion when it comes to who creates the metadata and the role of the publisher in coordinating this creation and distribution of metadata. Though the *Directory of Open Access Journals* (DOAJ) uses OA metadata standards, it does not create the metadata, as traditional, for-profit distributors would. Creating metadata requires specialized knowledge and a high level of collaboration and communication between library professionals, programmers, publishers, and distributors. This project investigates how OA journals in the DOAJ are using OA metadata standards for interoperability through coordinated efforts with librarians, programmers, and distributors and provide preliminary insight into how consistent and effective the methods are.

Background Information

Open Access Publishing, Economics and Ethics and Definitions

Traditionally content creation and distribution in scholarly publication has been tightly controlled. Academics review and approve articles for publication as part of their scholarship, which publishers and vendors distribute through subscriptions to individual journals or to databases (Suber, 2012; Brienza, 2011). The OA movement gained strength when the rise in prices for journal subscriptions and the decrease in library budgets increased the pressure to reduce obstacles to access to research (Dobson, 2003; Suber, 2012; Terry & Kiley, 2006; Shockey & Eisen, 2012). The OA movement also was strengthened when orders for monographs were reduced, further limiting publishing opportunities for scholars and the flow of the exchange of knowledge (Brienza, 2011; Fitzpatrick, 2011). Finally, scholars and librarians questioned the ethics of research funded by government or institutions being owned and controlled by for-profit publishers, and sought OA options as alternatives (Shockey & Eisen, 2012).

As a result, the benefits of building an OA system that is more equitable and offers more opportunity to engage in research to advance current knowledge allowed the OA movement to gain traction. In 2001, the *Budapest Open Access Initiative (BOAI)* defined the key elements of OA publications as "freely available," "online," "scholarly works" that are peer-reviewed and created as part of academic research, and licensed for free reuse with attribution (Bailey, 2006, p. 15). The *Berlin Declaration* and *The Bethesda Statement* adopted the *BOAI* guidelines, specifying the need for OA content to also be deposited in a repository to facilitate access and archiving (Bailey, 2006, pp. 17-18). Together, these formed what is known as the "BBB" definition of OA. OA journals vary from new publications using OA models to established publishers adding OA distribution to independent self-publishing models run by scholars and institutions (Bailey, 2006, pp. 24-25). However, management of that information remains an issue: "While open access theoretically provides greater and freer access to scholarly work, it can only do so if the material can be indexed in a way that people can find it and if the technology is made accessible to all" (Cheby, 2012, p. 4). Thus the question of who is responsible for the planning, management, and creation of metadata for individual journals remains a bit undefined.

Metadata and Open Access Metadata

Understanding and implementing OA metadata is essential for the success of OA publishing. In library science, "metadata is commonly used for any formal scheme of resource description, applying to any type of object, digital or non-digital" (*Understanding metadata*, 2004, p. 1). OA metadata is simply metadata that is "openly licensed and freely accessible," making any bibliographic metadata exposable and harvestable, whether the full content of the resources are OA or not (Flynn, 2013, p. 29). In that sense, OA metadata is broader in scope than OA publishing as it may be applicable to any publishing model. Since the function of metadata is primarily to facilitate the "discovery of relevant information," OA journals should use OA metadata to make their content discoverable in relevant searches (*Understanding metadata*, 2004, p. 1).

Interoperability

Interoperability is necessary for OA metadata to be used for discovery of content. Interoperability allows metadata to be read by "multiple systems with different hardware and software platforms, data structures, and interfaces ... with minimal loss of content or function" (*Understanding metadata*, 2004, p. 2). Interoperability may be achieved through *cross-system search* or through *metadata harvesting* (*Understanding metadata*, 2004, p. 2). A *cross-system search* maps searches to "a common set of attributes" (*Understanding metadata*, 2004, p. 2) rather than directly sharing metadata. *Metadata harvesting* requires providers produce OA metadata as "a common core set of elements" (*Understanding metadata*, 2004, p. 2) available for harvesting by a central index that may be searched by other repositories, databases, or libraries. As a result, OA metadata may appear in search results from

search engines like Google Scholar or from aggregate databases like EBSCOhost. This allows searchers to find the most relevant content regardless of where that content resides and is essential for OA content to be discoverable.

Composing, Coding, and Distributing Metadata

Metadata may be created by catalogers, who may be minimally trained or hold degrees in library and information science, working for any number of stakeholders, such as libraries, vendors, or publishers (Flynn, 2013, p. 30; *Understanding metadata*, 2004, p. 10). Metadata is composed by following a scheme. A scheme is any set of elements applied to a specific purpose, such as describing an object (*Understanding metadata*, 2004, p. 2). Each element in a scheme has a meaning, referred to as the semantics of the scheme, and the content of the metadata record is created by assigning a value to one or more of the elements (*Understanding metadata*, 2004, p. 2). For example, if an element titled 'creator' refers to the person or entity who created the object being described, then the value that should be assigned for that element is the name of that person or entity. Common schemes include Dublin Core (DC), Text Encoding Initiative (TEI), Metadata Encoding and Transmission Standard (METS), Metadata Object Description Schema (MODS), and Encoded Archival Description (EAD). Other schemes, like Learning Object Metadata (LOM), <indec>, and Online Information Exchange (ONIX) are specialized for certain types of media, such a learning objects, visual objects, or multimedia objects, or certain types of information, such copyright and attribution for ecommerce (*Understanding metadata*, 2004, pp. 3-8). Schemes are coded using computer-programming codes, such as Standard Generalized Markup Language (SGML), Extensible Markup Language (XML), and Hypertext Markup Language (HTML) (*Understanding metadata*, 2004, p. 2). This requires catalogers to know some programming, programmers to know some cataloging, or, ideally, for the specialists to work together.

Indexers or creators of metadata may use various software tools to assist in composing and coding the metadata, including templates, extraction tools, and conversion tools (*Understanding metadata*, 2004, p. 10). For templates, the information is entered by trained information professionals. Extraction and conversional tools are automated programs, but should be reviewed and edited by professionals since extraction and conversion are imprecise and subject to losing or incorrectly assigning values (*Understanding metadata*, 2004, p. 10). Thus, while tools may be helpful in speeding up the process, the human factor and cost cannot be completely eliminated.

OA metadata may be exposed or distributed through initiatives created by individual institutions, such as Harvard's OA Initiative and Online Computer Library Center's (OCLC) Open Data Commons Attribution License, or by housing OA metadata "in the cloud" from where it may be harvested or "pulled into local OPACs" (Flynn, 2013, p. 30). Either way, quality and interoperability of metadata is essential to fulfilling the mission of all OA initiatives and to the success of OA publishing.

Statement of the Problem

The unique traits of various types of content and the variety of missions of organizations and institutions that are indexing content calls for a variety of metadata schemes (*Understanding metadata*, 2004, p. 11). Thus, interoperability of metadata created for OA journals is essential for the content of these journals to be discoverable to users. A study by Cummings (2013) looked at the number of OA journals indexed in prominent databases, such as EBSCOhost, ProQuest, and Gale, but did not include information about who created the metadata and how it was exposed for harvesting. This project investigates how OA journals are applying emerging OA metadata standards and standards of interoperability in order to create quality metadata to make their content discoverable. In particular, the following questions are posed:

1. What metadata schemes and coding languages are OA journals using to create metadata?
2. Who creates this metadata for OA journals?
3. How do these practices compare with the best practices and standards for providing metadata for discoverability?

Methodology

Given the limited time and scope of this study, the researcher chose five journals in the subject area of library and information science from the DOAJ. Once each journal was identified, metadata records from two articles in each journal were examined for scheme choice as well as quality of metadata based on the inclusion of ample elements with assigned values for discoverability. The findings were compared to the standards or best practices proposed by prominent OA initiatives. The results provided preliminary insight into how OA journals create and provide OA metadata to make their content discoverable.

Literature Review

Lagace, Kaplan, and Leffler (2015) note "the creation of standards builds consensus within a community and facilitates interoperability among systems" and that standards start as recommended practices before becoming standards (p. 192). The emerging fields of OA metadata and publication are in the process of finding consensus among emerging recommended protocols. The literature on OA metadata and publishing covers the following areas: standards for OA metadata harvesting and the identification of OA content, metadata protocols based on the DC scheme and XML coding, best practices for creating OA collections, and practicalities of implementing these protocols and best practices.

Why We Need Better OA Metadata for Metadata Harvesting

Flynn (2013) defines OA metadata as "bibliographic information describing library content that is open licensed and freely accessible" (p. 29). While OA repositories

have flourished from one in 2003 to 250 by 2013, "the underlying infrastructure to support and sustain OA publishing" is just taking shape (Hodgson, 2014, p. 6). The community needs automated systems that are able to identify OA content regardless of where it is published and metadata standards to make this information readable by any system or by humans (Chumbe, Kelly & MacLeod, 2015; Graham, 2001; Hodgson, 2014). One example of insufficient standards is the lack of a metadata standard to clearly identify the licensing of content which results in OA articles in hybrid journals being overlooked by subscription services (Chumbe et al., 2015, pp. 143-144; Hodgson, 2014, p. 8). Another example is the use of PDFs that are not easily or accurately parsed by discovery systems for metadata and the lack of direct HTTP access to item information in online public access catalogs (OPAC), rendering that content undiscoverable by web crawlers or discovery systems (Graham, 2001, pp. 291-292). Metadata standards also should utilize unique identifiers for authors, institutions, and articles, such as DOI, in order to make OA content as discoverable as possible (Hodgson, 2014, p. 11). Creating OA metadata standards to make OA publications discoverable makes OA a more desirable form of publishing by allowing authors to publish in prestigious journals that may be made accessible via OA; it also allows web discovery services to discover OA content, which is currently hidden behind subscription walls (Chumbe et al., 2015, p. 145).

Recommended Protocols: OAI and NISO

Metadata harvesting protocols rely on the cooperation of two groups: data providers and service providers (Graham, 2001, p. 291). The data provider manages repositories or other systems that contain content and chooses the protocol, perhaps an OA metadata protocol, for exposing the metadata for items in their collection to harvesters (Lagoze, Van de Sompel & Nelson, 2002; Graham, 2001). A harvester issues OA metadata requests and is operated by service providers that deliver the item information to the end user (Lagoze et al., 2002; Graham, 2001). Open Archives Initiative (OAI) and NISO are the organizations leading the way in recommending protocols for OA metadata for metadata harvesting, though specialized groups have also created recommended protocols for metadata.

OAI's protocol, OAI - Protocol for Metadata Harvesting (OAI-PMH) "provides an application-independent interoperability framework based on *metadata harvesting*" (Lagoze et al., 2002, para. 1) in an attempt to create a low-cost standard that will make hidden content more discoverable (Graham, 2001, p. 291). OAI-PMH provides detailed specifications for creating and sharing OA metadata using the DC scheme and XML coding (Lagoze et al., 2002; Graham, 2001). The protocol supports sharing of metadata in multiple formats while requiring that each record use coding that identifies the metadata scheme, the URL for that scheme and the scheme's global identifier (Lagoze et al., 2002; Graham, 2001). However, to make the metadata interoperable, Dublin Core without qualifications must be used by repositories to share information (Lagoze et al., 2002; Graham, 2001; Efron, 2002). OAI-PMH also specifies the parts of the XML template that are required, including a header including unique identification

information and date, metadata describing the item in DC format without qualifications, and an optional code that provides data about the metadata using the XML scheme (Lagoze et al., 2002; Graham, 2001).

NISO provides four standards for information, one of which is the Open Access Metadata Indicators (OAMI) (Lagace et al., 2015, p. 191). OAMI is primarily focused on a metadata standard that indicates what level of open access the article provides, if any, and the copyright stipulations for the reader for that particular work (Lagace et al., 2015, p. 195). OAMI does not specify a scheme or coding, but proposes the inclusion of <free_to_read> tags or <license_ref> tags in order to indicate openness and the licensing of content (Lagace et al., 2015, p. 195). Under this protocol, the licensing tag must "include an Hyper Text Transfer Protocol (HTTP) uniform resource identifiers (URI) to point to license terms" (Lagace et al., 2015, p. 195) that are readable by machines and humans. An example might be directing the reader to a Creative Commons license. This recommended practice may be implemented and accessed by "readers, authors, publishers, funders, discovery services and search engines, and libraries" (Lagace et al., 2015, p. 195).

The Open Language Archives Community (OLAC) is an example of a specialized community that is working to standardize metadata for their own group. OLAC specifically requires "an XML format to interchange language-resource metadata within the framework of the Open Archives Initiative [OAI]" using all fifteen elements of DC (*OLAC Metadata*, n.d.). Similarly, the Journal TOCs: Expanding Market Opportunities (JEMO) project aims to expedite standardization of metadata by embedding strict OA elements in metadata, such as elements from DC and Creative Commons (CC), schemes publishers are already using (Chumbe et al., 2015, pp. 144-147).

Some find DC problematic and are not willing to specify it as a recommended protocol. Suber, from the Harvard Open Access Initiative, supports "adoption of community or discipline-specific *metadata vocabularies* that are more robust than Dublin Core" (Hodgson, 2014, p. 8). Efron (2007) calls Dublin Core "a rudimentary, weakly expressive standard in comparison to other archival metadata standards such as METS," ("Implications", para. 6) though he is not sure if more complex schemes are sustainable in for OA metadata interoperability. While not everyone likes DC, it is easy and cheap to convert from MARC, making it a strong contender among emerging recommendations (Graham, 2001, p. 293).

Best Practices for Creating Open Access

Statements of best practices for creating OA initiatives cover the drafting, adoption and implementation of policies governing requirements for OA publishing within an institution, but do not specify metadata schemes or formats (Scheiber & Suber, 2015; NISO Framework Working Group, 2007, p. 58-62). The Harvard Open Access Initiative's *Good Practices for University Open-Access Policies* specifies that indexing should allow items to be discoverable by search engines, but gives no scheme or coding specifications (Scheiber & Suber, 2015). NISO's *A Framework of Guidance for Building Good Digital Collections* provides six principles of

metadata, such as following community standards, interoperability, and indications of rights and licensing (NISO Framework Working Group, 2007, pp. 58-62). The Open Data Commons recommends that all data be made publicly available and have a license, such as the Open Data Commons Attribution (ODCA) (*Making your data open: A guide*, n.d.).

Implementing OA Metadata Standards

OA metadata is more likely to be implemented and accurate when it fits with schemes and coding that are already being used, when there is already an organized process and space for creating and storing metadata, and when there is motivation to use OA metadata (Chumbe et al., 2015, p. 150; Efron, 2007). For example, since "60% of publishers are already using DC elements" (Chumbe et al., 2015, p. 150) and licensing OA articles through Creative Commons (CC), it was not difficult to implement standards as suggested by the JEMO project. When the publishers create metadata in-house there is the "flexibility, skill, access and resources to modify the production systems" (Chumbe et al., 2015, p. 150) so that implementation may be quick and successful. Efron found that when catalogers use a strict structure for creating metadata and use multiple elements in their metadata schemes, information retrieval is better. Based on these two studies it is clear that buy-in from publishers by employing skilled catalogers is essential to successful implementation. Mandates from funders for OA publication and metadata also encourage implementation of OA protocols. Wellcome Trust requires OA publication and verifies through PubMed's central automated searches whether or not researchers are making their results available through OA; the results indicate approximately 70% compliance with some margin of error for false hits and missed articles (Hodgson, 2014, p.8).

Efron's (2007) study of how institutional repositories exposed metadata using OAI-PMH found that 19 out of 23 sampled repositories provided properly formed XML data and averaged more than 18 elements in each record. This shows improvement over an earlier study by Jewel Ward that showed an average of eight defined elements (Efron, 2007). Of the 65% of the records in the sample that used subjects, each often listed two or three subjects per record, a key element for appearing in relevant searches (Efron, 2007).

Obstacles to the implementation of OAI-PMH standards include incompatibility with word processing software used by content creators, the cost of programs that will easily and accurately convert the content to match the protocols, and the need for specialized knowledge to use tools designed for OA publishing and metadata creation, which is now in the hands of publishers rather than librarians (Hodgson, 2014, p. 12). There were four repositories in Efron's (2007) study that provided ill-formed XML with errors such as improper formatting or unpermitted characters in the code. Efron suggested two areas of improvement for OA metadata implementation: better XML formatting and proofreading, and an increase in the use of less common DC elements such as rights, coverage, and source to increase points for information retrieval.

Data Collection, Description, and Analysis

DOAJ Metadata Guidelines

DOAJ's website provides detailed expectations for how metadata should be formatted and provided to DOAJ. For both article and journal metadata DOAJ provides OAI-PMH feeds. DOAJ maps the OAI-PMH article metadata to DC elements according to definitions of each element as provided by the table on their website. DOAJ also lists additional DOAJ OAI fields – such as volume, issue, start and end pages – and definitions of each field to guide publishers who want to include such information in their metadata.

It is the publisher's responsibility to provide DOAJ with metadata for the journal and articles. Publishers may get article metadata into DOAJ by uploading an XML file or by completing a metadata information form for each article. If uploading, the DOAJ requires publishers convert their data to DOAJ XML file format. DOAJ provides a template for the XML scheme that may be edited in a basic text editor. This template includes elements that correspond to the DC scheme and the DOAJ-OAI fields. To assist publishers not fluent in coding languages, DOAJ also provides an example article record with explanations (see Figure 1) to make it easier to see where the specific journal and article information should be entered. For example, the right column points to where the publisher should replace the values of elements within the code, such as the ISSN number, author's name, and so on. The code could be copied and the element values edited. For instance, looking at this section of code:

```
<title language="eng">Roses and Lilies</title>
<authors>
  <author>
    <name>Fritz Haber
  </name>
```

The publisher may replace value for the title in the sample, "Roses and Lilies," with the title of the article being described in the metadata file. Likewise, the publisher may replace the author's name in the sample, "Fritz Haber," with the author's name in the article being described in the metadata file. Once the file is manually created, it may be uploaded to DOAJ.

<pre> <?xml version="1.0" encoding=" UTF-8"?> <records> <record> <language>eng</language> <publisher>Popular Chemist ry</publisher> <journalTitle>Botanical Ma gazine</journalTitle> <issn>17497221</issn> <eissn>17497234</eissn> <publicationDate>2002-09-3 0</publicationDate> <volume>98</volume> <issue>2</issue> <startPage>1234</startPage > <endPage>1235</endPage> <doi>1234567</doi> <publisherRecordId>12345</ publisherRecordId> <documentType>article</doc umentType> <title language="eng">Rose s and Lilies</title> <authors> <author> <name>Fritz Haber </na me> <email>fritz.haber@som e.university.org</email> <affiliationId>1</affi liationId> <affiliationId>2</affi liationId> <affiliationId>3</affi liationId> </author> </authors> <affiliationsList> <affiliationName affilia tionId="1"> University of A </affiliationName> <affiliationName affilia tionId="2"> University of B </affiliationName> <affiliationName affilia tionId="3"> University of C </affiliationName> </affiliationsList> </pre>	<pre> <= The language tag content m ust conform to the iso 639-2b standard. Find the correct language code. <= Here is the issn number of the journal in which the article has been publi shed. If you have an eissn number instead then this l ine should be replaced by 17497221. If you have both issn and eissn numbers then the eissn tag should occur after the issn tag. <= The author name should be f ormatted First Name, Middle Na me, Last Name <= Note that the affiliation Id numbers denote the affiliations in the affil iationslist below. </pre>
<pre> <abstract language="eng"> The catalytic formation of ammonia from hydrogen and atmospheric nitrogen under conditions of high temperature and high pre ssure. </abstract> <fullTextUrl format="pdf"> http://www.science.org/article s/HaberBosch.pdf </fullTextUrl> <keywords language="eng"> <keyword>garden</keyword > <keyword>rose</keyword> </keywords> </record> </record> ... </records> </pre>	<pre> <= Here is where your first r ecord ends. You may add as many more as you wa nt of similar kind below. </pre>

Figure 1. This figure presents the XML Sample Record image for the DOAJ recommended coding.

The metadata entry form, shown in Figure 2, allows publishers who are not familiar with coding or who cannot hire programmers to provide metadata without having to worry about errors in coding. Catalogers should recognize that the form contains the DC and DOAJ-OAI elements outlined in the XML instructions and may use the guides on the website to know what values to assign to each element to provide accurate and complete metadata.

The screenshot shows a web interface for entering article metadata. At the top, there are navigation tabs: "Your Journals", "Reapplications", "Upload Article XML", "Enter Article Metadata", and "Help". The "Enter Article Metadata" tab is active. The form contains the following fields and controls:

- Article Title ***: A text input field with the placeholder "Enter the article title here".
- DOI**: A text input field containing "10.1234/article-32".
- Author(s) ***: A section with three rows, each containing a "Name" input field and an "Affiliation" input field. Each row has a red "X" icon to its right. Below these rows is a blue button labeled "Add More Authors".
- Abstract**: A large text area with the placeholder "Enter the abstract here".
- Keywords**: A text input field with the placeholder "Enter some keywords for the article". Below it is the instruction "Use a , to separate keywords".
- Full-Text URL**: A text input field with the placeholder "URL to the fulltext of the article". Below it is the instruction "(The URL for each article must be unique)".
- Publication Date**: A dropdown menu showing "07" and "2015".
- Journal ISSN (print version) ***: A dropdown menu.
- Journal ISSN (online version) ***: A dropdown menu.
- Journal**: A dropdown menu.
- Volume**: A text input field with the placeholder "Vol".
- Issue Number**: A text input field with the placeholder "Issue".
- Page(s)**: Two text input fields, "Start" and "End", with the placeholder "Start" and "End" respectively.
- Add Article**: A green button at the bottom of the form.

Figure 2. This figure presents a screenshot of DOAJ's publisher's metadata entry form.

Choosing Journals & Retrieving Metadata

Five journals from a subject search for Library and Information Science in the DOAJ search tool were randomly chosen:

- *Informing Science The International Journal of an Emerging Transdiscipline*
- *Code4Lib Journal*
- *Journal of Library Innovation*
- *In the Library with the Lead Pipe*
- *Journal of Librarianship and Scholarly Communication*

Two articles from each journal's main website were selected for OA metadata retrieval and examination.

DOAJ makes metadata records available to any OAI compatible service. Since the researcher does not have access to an OAI compatible service, Ann Agee, Librarian for the School of Information at San Jose State University (SJSU) and Steven Higaki, the Head of Cataloging for the King Library at SJSU, were contacted to see if the university could request the metadata. Neither has access to OA metadata files. Higaki explained the King Library does not harvest metadata directly from DOAJ, "The catalog records you see are part of a service we subscribe to that assists with the management of our electronic resources/journals" (personal communication, July 23, 2015). Both Agee and Higaki suggested using the page source code or contacting the publishers directly (personal communication, July 23, 2015). Though WorldCat contains metadata, a permission code is needed to access them.

Metadata for two articles from each journal was requested directly from DOAJ and each journal along with information about which method was used to provide the metadata to DOAJ. The links to metadata provided by *Journal of Librarianship and Scholarly Communication (JLSC)* resulted in an error code and *In the Library with the Lead Pipe* did not respond, eliminating both from the study. Of the three remaining journals, one does not yet provide metadata to DOAJ, one uses the DOAJ publisher's form, and one provides metadata through XML files. For each article chosen a title search in the DOAJ search tool was conducted. The journal that did not provide metadata did not have article level search results in the DOAJ, though all were discoverable by title in Google Scholar. Both the journal that provided the XML file to DOAJ and the one that used the online form to provide metadata were discoverable by title using the DOAJ search tool. The journals' responses to how they provide metadata, the articles to be examined, and the availability of article level information in the DOAJ search tool are presented in Table 1.

Table 1			
<i>Metadata Provided and Article Findability by Title Search</i>			
Journal	Article	Title Found in DOAJ?	Format Metadata Provided
<i>Informing Science The International Journal of an Emerging Transdiscipline</i>	Information Gatekeepers – Aren't We All?	No	None ¹
<i>Informing Science The International Journal of an Emerging Transdiscipline</i>	The Impact Facebook and Twitter has on the Cognitive Social Capital of University Students	No	
<i>Journal of Library Innovation</i>	Addressing Rural Library Technology Budgets with Single Board Computers: Testing the APC 8950 Rock Circuit Board Computer for Patron Access	Yes	DOAJ Form ²
<i>Journal of Library Innovation</i>	Open Education Resources: The New Paradigm in Academic Libraries	Yes	
<i>Code4Lib Journal</i>	Recognizing Cultural Diversity in Library Interface Development	Yes	XML Upload ³
<i>Code4Lib Journal</i>	“What If I Break It?”: Project Management for Intergenerational Library Teams Creating Non-MARC Metadata	Yes	

Notes:

¹E. Cohen, personal communication, July 22, 2015.

²D. Schoen, personal communication, July 24, 2015

³S. Amato, personal communication, July 22, 2015

This discrepancy in title searches in DOAJ and Google Scholar implied that information might be available in the HTML coding even if the OA metadata is not coded by DOAJ standards and provided to DOAJ. Since all were not able to provide metadata files, in addition to XML or other metadata file provided by the journal, the following data was also gathered for each article, as available, in order

to see if any of the DOAJ metadata protocols were being included in the source code for articles and if the source code, which is what Google Scholar would search, for each article is different than the metadata provided to DOAJ:

- Source code from the article page on the journal website using a Safari browser,
- Source code from the article page on DOAJ using a Safari browser.

Analysis of Article Metadata

Each article's source code was examined for required DOAJ elements as DC, XML, or XSD scheme and values and compared to the XSD template and example record. This allowed for a sense of how successfully the shared metadata conveyed the needed information for discovery of the articles by researchers.

The code from *Journal of Library Innovation* clearly contains XML and DC metadata within the source code from the website. The following tables show excerpts of the source code next to the DOAJ XSD file code. Overall, the source code of each article contains the same elements as the XSD file and contains explicit DC code to map with DOAJ. In Table 2 it can be seen that all the codes begin with the identical XML line, however, rather than mapping to the W3C XMLScheme, the page code maps to the W3C XHTML.

In Table 3, the source code excerpted shows how the HTML code includes meta tags for elements such as the article and journal title that have no parallel in the XSD file. As seen in Table 4, this information repeats in code that does follow the DC scheme.

Article 1 Source Code	Article 2 Source Code	DOAJ XSD file
<pre><?xml version="1.0" encoding="UTF-8"?> <!DOCTYPE html PUBLIC "- //W3C//DTD XHTML 1.0 Transitional//EN" "http://www.w 3.org/TR/xhtml1/D D/xhtml1- transitional.dtd"> <html xmlns="http://www.w 3.org/1999/xhtml"> <head></pre>	<pre><?xml version="1.0" encoding="UTF-8"?> <!DOCTYPE html PUBLIC "- //W3C//DTD XHTML 1.0 Transitional//EN" "http://www.w 3.org/TR/xhtml1/D D/xhtml1- transitional.dtd"> <html xmlns="http://www.w 3.org/1999/xhtml"> <head></pre>	<pre><?xml version="1.0" encoding="UTF-8"?> <xs:schema xmlns:xs ="http://www.w3.org/ 2001/XMLSchema" xmlns:iso_639- 2b="http://www.doaj. org/schemas/iso_639- 2b/1.0"></pre>

Table 3		
<i>Excerpts from Source Code for Journal of Library Innovation</i>		
Article 1 Source Code	Article 2 Source Code	DOAJ XSD file
<pre><meta http-equiv="Content-Type" content="text/html; charset=utf-8" /> <meta name="description" content="Open Education Resources: The New Paradigm in Academic Libraries" /> <meta name="keywords" content="Scholarly Communications; Open Scholarship; Alternative Educational Materials; Open Access; Open Educational Resources; Institutional Repositories" /></pre>	<pre><meta http-equiv="Content-Type" content="text/html; charset=utf-8" /> <meta name="description" content="Addressing Rural Library Technology Budgets with Single Board Computers: Testing the APC 8950 Rock Circuit Board Computer for Patron Access" /> <meta name="keywords" content="Single board computers; Raspberry Pi; rural libraries; linux; android; budgets; technology" /></pre>	No equivalent code in this file.

Table 4 shows where the source code in both articles includes a reference to the DC scheme. The XSD file does not show this, though the DOAJ does map to DC. The inclusion of DC elements in all three codes is seen in the rest of Table 4.

Excerpts of common points where the DC and XSD may map smoothly, such as the fields for language, title, descriptions, ISSN, are included in Table 4. Thus, in the journal's source code for each article, metadata tags that duplicate the DC elements comingle with html coding that makes the content visible on journal's webpage. Since this journal included DC coding in the source code, it is interesting to note that the publisher used the DOAJ form to provide the metadata to the DOAJ rather than excerpting that portion of the code to send in XML format.

Table 4		
<i>Excerpts of Source Code for Journal of Library Innovation</i>		
Article 1 Source Code	Article 2 Source Code	DOAJ XSD file
<pre><link rel="schema.DC" href="http://purl.org/dc/elements/1.1/" /> name="DC.Language" " scheme="ISO639-1" content="en"/> <meta name="DC.Source" content="Journal of Library Innovation"/> <meta name="DC.Source.IS SN" content="1947- 525X"/> <meta name="DC.Source.Iss ue" content="1"/> <meta name="DC.Source.U RI" content="http://www.l ibraryinnovation.org/" /> <meta name="DC.Source.Vo lume" content="5"/> <meta name="DC.Subject" xml:lang="en" content="Scholarly Communications"/></pre>	<pre><link rel="schema.DC" href="http://purl.org/dc/elements/1.1/" /> ... <meta name="DC.Language" " scheme="ISO639-1" content="en"/> <meta name="DC.Source" content="Journal of Library Innovation"/> <meta name="DC.Source.IS SN" content="1947- 525X"/> <meta name="DC.Source.Iss ue" content="1"/> <meta name="DC.Source.U RI" content="http://www.l ibraryinnovation.org/" /> <meta name="DC.Source.Vo lume" content="5"/> <meta name="DC.Subject" xml:lang="en" content="Single board computers"/></pre>	<pre><xs:element name="language" type="iso_639- 2b:LanguageCodeTyp e" minOccurs="0"/> <xs:element name="publisher" type="xs:string" minOccurs="0"/> <xs:element name="journalTitle" type="xs:string" /> <xs:element name="volume" type="xs:string" minOccurs="0"/> <xs:element name="issue" type="xs:string" minOccurs="0"/> <xs:element name="startPage" type="xs:string" minOccurs="0"/> <xs:element name="endPage" type="xs:string" minOccurs="0"/></pre>

The source code for the articles from *Code4Lib* does not start with the XSD code of the sample file. Even though the journal that provided they XML file to DOAJ, they did not include the XSD coding in the source code as seen in the source code from the *Journal of Library Innovation* in Table 4. Like the *Journal of Library Innovation*, rather than mapping to the W3C XMLSchema, the page code maps to the W3C XHTML scheme for public presentation, again, likely because this is code for the actual article page on the journal's website.

Table 5		
<i>Excerpts of Source Code for Code4Lib</i>		
Article 1 Source Code	Article 2 Source Code	DOAJ XSD file
<pre><!DOCTYPE html PUBLIC "- //W3C//DTD XHTML 1.0 Transitional//EN" "http://www.w3.org/T R/xhtml1/DTD/xhtml1 -transitional.dtd"> <html xmlns="http://www.w 3.org/1999/xhtml" lang="en-US"> <head> <meta http- equiv="Content-Type" content="text/html; charset=UTF-8" /> <title>The Code4Lib Journal &#8211; Recognizing Cultural Diversity in Library Interface Development</title></pre>	<pre><!DOCTYPE html PUBLIC "- //W3C//DTD XHTML 1.0 Transitional//EN" "http://www.w3.org/T R/xhtml1/DTD/xhtml1 -transitional.dtd"> <html xmlns="http://www.w 3.org/1999/xhtml" lang="en-US"> <head> <meta http- equiv="Content-Type" content="text/html; charset=UTF-8" /> <title>The Code4Lib Journal &#8211; “What If I Break It?”: Project Management for Intergenerational Library Teams Creating Non-MARC Metadata</title></pre>	<pre><?xml version="1.0" encoding="UTF-8"?> <xs:schema xmlns:xs ="http://www.w3.org/ 2001/XMLSchema" xmlns:iso_639- 2b="http://www.doaj. org/schemas/iso_639- 2b/1.0"></pre>

Though the elements, such as the article title, the author, and the ISSN, required by the DOAJ schemes are visible in the code from *Code4Lib*, these items are not coded using the DC scheme, as seen in Table 6. Therefore, it is unclear if these values will successfully map according to OAI-PMH standards for metadata. Since this journal provided XML files to DOAJ, it is surprising to not find this scheme included in the source code. Instead, it appears the journal chose to code

two separate metadata files, one for the website and one to provide the metadata to DOAJ according to their specifications.

Table 6	
<i>Excerpts of Source Code for Code4Lib</i>	
Article 1 Source Code	Article 2 Source Code
<pre><div class="article" id="post-10456"> <p id="issueDesignation">Issue 28, 2015-04- 15</p> <h1 class="articletitle">Recognizing Cultural Diversity in Library Interface Development</h1> <div class="abstract"> <p>The rapid increase in complex library digital infrastructures has enabled a more full-featured set of resources to become accessible by autonomous users,...exploring the redevelopment strategy for the New York University Libraries' web presence, which serves a broad and global set of users.</p> </div> <div class="entry"> <abbr class="unapi-id" title="http://journal.code4lib.org/?p=1 0456"><!-- &nbsp; --></abbr> <p>by Nik Dragovic</p></pre>	<pre><div class="article" id="post-10395"> <p id="issueDesignation">Issue 28, 2015-04- 15</p> <h1 class="articletitle">"What If I Break It?": Project Management for Intergenerational Library Teams Creating Non-MARC Metadata</h1> <div class="abstract"> <p>Libraries are constantly challenged to meet new user needs and to provide access to new types of materials. ...orient themselves when embedded in a "traditional" library setting.</p> </div> <div class="entry"> <abbr class="unapi-id" title="http://journal.code4lib.org/?p=1 0395"><!-- &nbsp; --></abbr> <p>by Kelly J. Thompson</p></pre>

The source code from *Informing Science*, presented in Table 7, contains more lines with metadata with DC elements than that *Code4Lib*, though it also does not follow the DC scheme or the XML requirements for the DOAJ. Since this journal did not provide any metadata to the DOAJ, it is not surprising to find the DC and XML elements missing from their source code.

Table 7	
<i>Excerpts from Source Code for Informing Science Articles</i>	
Source Code for Article 1	Source Code for Article 2
<pre> <!DOCTYPE html> <!--[if IE 8]> <html class="no-js lt-ie9" lang="en"> <![endif]--> <!--[if gt IE 8]><!--> <html class="no-js" lang="en"> <!--<![endif]--> <head> <meta charset="utf-8" /> <meta http-equiv="X-UA- Compatible" content="IE=edge" > <meta name="description" content="An international association advancing the multidisciplinary study of informing systems. Founded in 1998, the Informing Science Institute (ISI) is a global community of academics shaping the future of informing science."> <meta name="viewport" content="width=device-width" /> <title>Informing Science Institute - Information Gatekeepers – Aren't We All?</title> <link rel="shortcut icon" href="/favicon.ico"> </pre>	<pre> <!DOCTYPE html> <!--[if IE 8]> <html class="no-js lt-ie9" lang="en"> <![endif]--> <!--[if gt IE 8]><!--> <html class="no-js" lang="en"> <!--<![endif]--> <head> <meta charset="utf-8" /> <meta http-equiv="X-UA- Compatible" content="IE=edge" > <meta name="description" content="An international association advancing the multidisciplinary study of informing systems. Founded in 1998, the Informing Science Institute (ISI) is a global community of academics shaping the future of informing science."> <meta name="viewport" content="width=device-width" /> <title>Informing Science Institute - The Impact Facebook and Twitter has on the Cognitive Social Capital of University Students</title> <link rel="shortcut icon" href="/favicon.ico"> </pre>

Table 8		
<i>Excerpts of Source Code from DOAJ for Code4Lib and Journal of Library Innovation Articles</i>		
Journal	Source Code for Article 1	Source Code for Article 2
Code4Lib	<pre><meta name="citation_journal_title" content="Code4Lib Journal"> <meta name="citation_publisher" content="Code4Lib"> <meta name="citation_author" content="Nik Dragovic"> <meta name="citation_title" content="Recognizing Cultural Diversity in Library Interface Development"> <meta name="citation_publication_date" content="2015/04/01"> <meta name="citation_issue" content="28"> <meta name="citation_issn" content="1940-5758"></pre>	<pre><meta name="citation_journal_title" content="Code4Lib Journal"> <meta name="citation_publisher" content="Code4Lib"> <meta name="citation_author" content="Kelly Thompson"> <meta name="citation_title" content="“What If I Break It?”: Project Management for Intergenerational Library Teams Creating Non-MARC Metadata"> <meta name="citation_publication_date" content="2015/04/01"> <meta name="citation_issue" content="28"> <meta name="citation_issn" content="1940-5758"></pre>
Journal of Library Innovation	<pre><meta name="citation_journal_title" content="Journal of Library Innovation"> <meta name="citation_publisher" content="Western New York Library Resources Council"> <meta name="citation_author" content="Michael D. Wells"> <meta name="citation_title" content="Addressing Rural Library Technology Budgets with Single Board Computers: Testing the APC 8950 Rock Circuit Board Computer for Patron Access"> <meta name="citation_publication_date" content="2014/04/01"> <meta name="citation_volume" content="5"> <meta name="citation_issue" content="4"> <meta name="citation_firstpage" content="1"> <meta name="citation_lastpage" content="12"> <meta name="citation_issn" content="1947-525X"></pre>	<pre><meta name="citation_journal_title" content="Journal of Library Innovation"> <meta name="citation_publisher" content="Western New York Library Resources Council"> <meta name="citation_author" content="Carmen Mitchell"> <meta name="citation_author" content="Melanie Chu"> <meta name="citation_title" content="Open Education Resources: The New Paradigm in Academic Libraries"> <meta name="citation_publication_date" content="2014/04/01"> <meta name="citation_volume" content="5"> <meta name="citation_issue" content="1"> <meta name="citation_firstpage" content="13"> <meta name="citation_lastpage" content="29"> <meta name="citation_issn" content="1947-525X"></pre>

The source code from the articles from *Code4Lib* and *Journal of Library Innovation* that appear on the DOAJ website are more uniform and reflect the DOAJ protocols. All begin with the same html code header:

```
<!DOCTYPE html>
<html dir="ltr" lang="en">
<head>
  <meta charset="utf-8">
```

The fourth line directs any metadata to the same encoding language used in the first line of the sample XSD file "<?xml version="1.0" encoding="UTF-8"?>." Further down in the file, as presented in Table 8, we can see metadata information that corresponds to the elements required by DC and included in DOAJ's XSD sample. Though not in DC scheme, it contains the elements and the XML is able to be mapped to DC scheme. For example, the elements of "citation_journal_title", "citation_publisher", and "citation_author" have equivalent elements in DC to which they may be easily mapped. Notice these files do not have the extra HTML tags that were seen in the source code from the journals' websites.

The XML file provided by *Code4Lib* looks identical to the example XML file provided on the DOAJ website as seen in Figure 1 above. Here is the first record in that file:

```
<?xml version="1.0" encoding="UTF-8"?>
<records>
  <!--
    Generated by the DOAJ Export WordPress plugin.
    http://wordpress.org/extend/plugins/doaj-export/
  -->
  <record>
    <language>eng</language>
    <publisher>Code4Lib</publisher>
    <journalTitle>The Code4Lib Journal</journalTitle>
    <issn>19405758</issn>
    <publicationDate>2015-07-15</publicationDate>
    <issue>29</issue>
    <publisherRecordId>10796</publisherRecordId>
    <documentType>article</documentType>
    <title language="eng">Editorial Introduction: Changes on the Editorial
    Board</title>

    <authors>
      <author>
        <name>Sara Amato</name>
      </author>
    </authors>
    <abstract language="eng">The publication of the 29th issue of the
    journal brings with it several changes to the editorial board.</abstract>
```

```

<fullTextUrl
format="html">http://journal.code4lib.org/articles/10796</fullTextUrl>
</record>

```

Thus it is clear this journal consulted and followed the DOAJ specifications for metadata creation and specified this as a part of the process for publishing as an OA publication.

Though *Code4Lib* is the only one that creates XML files, *Journal of Library Innovation* has the most metadata information in their source code on their website. When looking at the source code for the article from both journals on the DOAJ website, the metadata information seems to be equal. *Informing Science* does not provide metadata and its articles are not accessible on the website since no metadata information was uploaded so only the website source code was available.

Conclusions

A few obstacles to creating complete OA metadata for journals in DOAJ are implied by the information gathered in this project. Based on the lack of consistency of the source codes, not to be confused with accuracy, which was not part of the scope of this investigation, and communications with the journals, further studies might seek to confirm these obstacles to creating accurate and interoperable metadata:

1. the awareness by publishers of the standards and options to get metadata into DOAJ despite both being on the website;
2. the level of technical knowledge about coding and cataloging of publishers and its relation to their ability to implement the guidelines on the DOAJ site or to the option to use the form to enter metadata information;
3. the cost of personnel, especially for OA journals, with expertise in cataloging and/or programming to either manually enter metadata information for each article into the DOAJ website or to create xml files to upload to DOAJ.

Additionally, the impact of discrepancies in metadata creation on harvesting of metadata and discoverability of individual articles is an area that needs further research. If the discoverability does not correlate to the journal providing the metadata to DOAJ, then how are these OA journals, or any article, being indexed to be discovered in these searches? While much research has been done around the ethics and economics of OA publishing and indexing, there are not many studies about the creation, costs, and effects of OA metadata in relation to OA publication and discoverability. Moreover, with the emergence of hybrid journals in which some articles are published as OA and others not, and traditional journals publishing under modified forms of OA, these questions about OA metadata concern more than just OA journals.

While standards for metadata in the many areas of library and information sciences are still emerging, a forerunner for OA metadata standards seems to be the OAI-PMH, using XML and DC, as outlined in the DOAJ guidelines. The analysis

of the source codes gathered from the article pages on the journals' websites and from the article pages on DOAJ, and of metadata files provided by DOAJ or the journals, shows that there are multiple ways to include and share OA metadata. Though all have overlapping elements containing bibliographic information, there does not seem to be much consistency from journal to journal. Moreover, OA metadata is not explicitly created and shared, it could render articles invisible in some searches even if the elements of emerging schemes like DC are included. This is seen in the *Informing Science* articles that contain DC elements, but do not use the DC scheme or code in XML to be shareable with the DOAJ, resulting in the articles not being discoverable in a title search on DOAJ. There seemed to be little difference in the ability to do a title search for the other two journals' articles, though they included different elements in their source code and provided the OA metadata to DOAJ using different methods.

Since the creation of OA metadata was found to be the responsibility of the publishers, publishers must include a plan for creating and sharing metadata in order to ensure the discoverability of their journal's content. For any journal seeking to implement a plan for creating OA metadata, using XML and DC as outlined in the DOAJ guidelines and following OAI-PMH protocols are viable and effective options. However, many publications seem unaware of this. As more journals move to some form of OA publication, there needs to be more communication and collaboration in creating OA metadata in order to ensure that content is consistently discoverable. One way to do this is to foster collaboration between librarians and programmers, each who have specialized skills required to provide accurate content for the elements in the metadata scheme and to properly program these schemes into code that is readable by machines and humans. Though further study is needed to determine if greater standardization would help this process or if is adequate for each journal to have their own method as long as they are using scheme and coding that is interoperable.

References

- Bailey, C. W. (2006). What is open access? In N. Jacobs (Ed.), *Open access: Key strategic, technical and economic aspects* (pp. 13-26). Oxford: Chandos House.
- Brienza, C. (2012). Opening the wrong gate? The academic spring and scholarly publishing in the humanities and social sciences. *Publishing Research Quarterly*, 28(3), 159-171.
- Cheby, L. (2012). *Open Access: Revolution or Restructure?* (Unpublished student paper). San Jose State University, San Jose, CA.
- Chumbe, S., Kelly, B., & MacLeod, R. (2015). Hybrid journals: Ensuring systematic and standard discoverability of the latest open access articles. *The Serials Librarian: From the Printed Page to the Digital Age*, 68(1-4), 143-155. doi:10.1080/0361526X.2015.1016856

- Cummings, J. (2013). Open access journal content found in commercial full-text aggregation databases and journal citation reports. *New Library World*, 114(3/4), 166-178. doi:10.1108/03074801311304078
- Dobson, C. (2003). From bright idea to beta test: the story of LOCKSS. *Searcher*, 11(2), 51-53.
- Efron, M. (2007). Metadata use in OAI-compliant institutional repositories. *Journal of Digital Information*, 8(2). Retrieved from <https://journals.tdl.org/jodi/index.php/jodi/article/view/196/169>
- Fitzpatrick, K. (2011). *Planned obsolescence: publishing, technology, and the future of the academy*. New York: New York University Press. Print.
- Flynn, E. A. (2013). Open access metadata, catalogers, and vendors: The future of cataloging records. *The Journal of Academic Librarianship*, 39(1), 29-31.
- Graham, R. A. (2001). Metadata harvesting. *Library Hi Tech*, 19(3), 290-295. Retrieved from <http://search.proquest.com/docview/200596980?accountid=130717>
- Hodgson, C. (2014) Open access infrastructure: where we are and where we need to go. *Information Standards Quarterly*, 26(4-6), 4-14. doi: 10.3789/isqv26no2.2014.02
- Lagace, N., Kaplan, L., & Leffler, J. (2015). Actions and updates on the standards and best practices front. *The Serials Librarian: From the Printed Page to the Digital Age*, 68(1-4), 191-196. doi:10.1080/0361526X.2015.1017420
- Making your data open: A guide. (n.d.) *Open Data Commons*. Retrieved from <http://opendatacommons.org/guide/>
- NISO Framework Working Group. (2007). *A framework of guidance for building good digital collections*. Baltimore, MD: National Information Standards Organization.
- ODC attribution summary. (n.d.). *Open Data Commons*. Retrieved from <http://opendatacommons.org/licenses/by/summary/>
- OLAC metadata. (n.d.). *Open Language Archives Community*. Retrieved from <http://www.language-archives.org/OLAC/metadata.html>
- Reich, V. (2008). CLOCKSS—It Takes a Community. *The Serials Librarian*, 54:1-2, 135-139.
- Sheiber, S., & Suber, P. (2015). Good practices for university open-access policies. Retrieved from http://cyber.law.harvard.edu/hoap/Good_practices_for_university_open-access_policies
- Shockey, N. (Narration), & Eisen, J. (Narration) (2012). *Open access explained!* [Video file]. Retrieved from <http://www.youtube.com/watch?v=L5rVH1KGBCY>

- Suber, P. (2012). "What is open access?" *Open access*, MIT Press. Retrieved from http://cyber.law.harvard.edu/hoap/Open_Access
- Terry, R., & Kiley, R. (2006). Open access to the research literature: a funder's perspective. In N. Jacobs (Ed.), *Open access: Key strategic, technical and economic aspects* (pp. 101-109). Oxford: Chandos House.
- Understanding metadata*. (2004). Bethesda, MD: NISO Press.