Coding Discovery Between 3 Systems: Omeka, Alma & Primo

Keiko Okuhara Metadata Services Librarian

University of Hawai'i at Mānoa, William S. Richardson School of Law

Introduction

This session will introduce a sustainable process for promoting faculty publications through an online catalog by making three systems interoperable; Omeka (Data provider), Alma (Library System), and Primo (Discovery platform) by devising the version 2.0 Open Archive Initiative Protocol for Metadata Harvesting (OAI-PMH) in other words, OAI protocol. The OAI-PMH Repository plugin of the Omeka Classic allows an institutional repository (IR), a data provider, to be harvested in the discovery platform, Primo, to make a bridge for repository interoperability.

Omeka and its plugin Features

Omeka is an open-source digital publishing platform to organize digital collections, archives, oral histories, and essays as online exhibits. The basic functions of Omeka are uploading digital objects, adding metadata, and organizing objects into collections or exhibits for public display. Omeka can be used to exhibit almost any kind of digital object: image, sounds, video, text such as a portable documentation format (PDF) file or others with a 2 MB size limit per file. The item in Omeka may include multiple files. These files are described by Doblin Core with elements customizable by the item types. They are types of object such as moving and still images, email, lesson plan, oral history, person, Interactive resource, Website, dataset, etc. Visual elements can be customized, including colors, design, fonts, and header images. The plugin features enable Omeka to connect to other sites to upload audio files to SoundCloud, import images from Flicker, or auto-post to Tumblr in addition to adding visual enhancements like slideshow carousels, simple web pages for more extensive exhibit creation. Other plugins provide additional metadata-related feature like controlled vocabulary enhancements, metadata harvesters data to allow saving items to Zetero, or Dublin Core extended properties.

A Plugin feature is a software add-on that is installed on a software to enhance its capabilities. Plugins are a way to make functionality easy to maintain. The plugin features have been substantially redesigned to be more useful throughout the platform. The plugin features in Omeka roughly have the following seven types: 1) Library Administration, 2) Customization, 3) Integration Viewer, 4) Information Searching, 5) Data Export/Import, 6) Media File and 7) Miscellaneous.

The library administration plugin works for Institutional Repository Metadata. Its functions are to upload all authority files and subject headings from the Library of Congress such as LC Suggest, VRC Core, Simple Vocab. In the customization plugins, there are Admin Images, CSS Editor, IIIIF Toolkit, Posters, Select 2, Simple Contact Form, Simple Pages for designing the Home page or Interface page. The integration viewer plugins are to integrate document viewer with Omeka such as Docs Viewer, HTML 5 Media, PDF Embed, Universal viewer. The information searching plugins are to locate specific information or metadata in documents. They are Avant Search and Avant Common. The Data Export & Import plugins are an automated or semi-automated output and input data transfer feature from one application to another. They are Batch Uploader, CSV Export Format, CSV Import , Dropbox, Flicker Import, Import METS Export, and Omeka API import. The Media file plugins are the communication tools to store information or data to allow Omeka administrators to upload images not attached to items and to create structure metadata for images and videos. The maintenance plugins in Omeka can make a link to a document on Google Drive or any other type of document accessible through a URL.

Making a Bridge between the Institutional Repositories (IR) and Library Online Catalog/Discovery Platform

As a small or medium-sized institution, the William S. Richardson School of Law at the University of Hawai'i at Mānoa (the Library) implemented the Omeka digital platform, which has a built-in search tool, for its Institutional Repository. Although a state of art search engine wasn't built in the digital platform, Omeka is user-friendly, flexible, and collaborative with many plugins to be interoperable. The goal of the integration project was to make our Archival/Institutional collections more visible and accessible. All along, the Library was looking into the possibility of building an open source platform such as blacklight, but in-house staffing and resources were limited. Also, the Library could not afford the Digital Commons platform due to its high implementation and subscription costs. Then the necessity of utilizing existing resources including staffing emerged.

The author suggested the idea of making a bridge between Omeka and Alma/Primo as an interface in a sustainable fashion to employ a state-of-the-art search engine in the Primo discovery platform for greater searching results to make the library online catalog vital and to leverage the IR for more exposure of its contents. The bridge between different platforms breaks down barriers and allows users to augment the search results with fewer searches and less time. The enriched online catalog provides users with a greater variety of materials and makes it as a robust one-stop shopping tool to locate different kinds of content and library resources.

A discovery platform such as Primo can search library materials and items and records of the library materials hosted in the integrated library system (ILS) as a mega-aggregated knowledge database. The discovery platform has a single index of metadata from the various

library technology system to provide discovery of the library resources, both local and licensed through the discovery service. The local materials indexed in the discovery platform enable these materials findable. They include but not limited to archival materials, digital collections, institutional repository contents, and information uploaded on the library website such as LibGuides.

When metadata harvested by the discovery platform through many different library technology systems as well as a discovery services' central index, these metadata will be used in different ways for different things inside the platform. Metadata in various formats from diversified repositories may require transformation into another metadata format before it can be indexed, searched, and displayed with in the discovery platform. Ex Libris's Primo discovery platform has to be transformed or mapped into the Primo Normalized XML (PNX[1]) This transformation would occur inside of the platform or within the repository before the ingestion of the metadata.

While the discovery platform can harvest metadata from various library databases and repositories, the other method to deposit metadata from an institutional repository and digital library is to bulk export the metadata to load the metadata records in bulk into the discovery platform, which typically requires a fair amount of manual processing and staff time. The advantage of automatic harvesting of metadata is to aggregate records from multiple sources and to transmit it into one system, the discovery platform. The most commonly used method of harvesting local metadata is the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH), or OAI protocol. OAI-PMH was created by the Open Archives Initiative and was designed to be a versatile and simple protocol for harvesting metadata. It is based on HTTP and XML to "facilitate the aggregation and distribution of metadata" (Sandarson, Young, and LeVan, 2005) with great compatibility in format. OAI-PMH allows for metadata from various sources to be aggregated into one central database and can harvest any metadata in any schema written in eXtensible Markup Language (XML) to ensure a basic level of interoperability. OAI-PMH requires that metadata be in unqualified Doblin Core in addition to any other metadata formats. In this way, OAI-PMH allows the creation of central indexes within the discovery platform and makes harvested metadata searchable and accessible through the platform.

One of the limited features in OAI-PMH is that it isn't "designed to be used for the sharing of the actual digital objects that are contained in a repository." (Corrado, 2018) This limitation becomes an issue in exchanging or harvesting the actual digital content. The Open Archives Initiative Object Reuse and Exchange (OAI-ORE) is intended to "expose the rich content in these aggregations to applications that support authoring, deposit, exchange, visualization, reuse, and preservation" (https://www.openarchives.org/ore para.1) OAI-PMH

[1] PNX Record Section accessed December 10, 2021 from https://knowledge.exlibrisgroup.com/Primo/Product_Documentation/Primo/Technical_Guide/010The_PNX_Record/010PNX_Record_Sections

also maintains the ResourceSync Framework Specification[2] (ANSI/NISO Z39.99-2017[3]) to "describe a synchronization framework for the web consisting of various capabilities that allow third-party systems to remain synchronized with a servers' evolving resources" (NISO & Open Archies Initiative, 2017, Abstract) (http://www.openarchives.org.rs.toc)

Reflection and Next Steps

The integration project opens up a Pandora's box when the online library catalog is populated with content hosted in Omeka digital publishing platform. It brings excitement and innovation in exploring the possibility to improve the quality of the harvested metadata. It is an unconventional step to integrate IR content into the library catalog by automating metadata harvesting by way of a Uniform Resource Locator (URL), http(s) with the OAI-PMH initiation that transforms Dublin Core in IR into Extensible Markup Language (XML) to enhance IR discoverability. Here are the findings from the case of the University of San Diego. It is experimented that they "learned that they would have greater flexibility in customizing the output of XML data from the IR by utilizing the metadata prefixes dcs (simple-dublin-core) or dcq (qualified-dublin-core) rather than oai-dc in the harvesting formula." (Makula 204) If this theory applies to the OAI-PMH harvesting in general, it's worth investigating for the Library to analyze how the harvested metadata are mapped into the Alma and how these metadata are transferred and displayed on Primo, discovery platform, since Alma supports only oai_dc at this point. If the default metadata prefix is set as "oai_dc," the Dublin Core elements are pre-mapped to certain fields, and the pre-mapped metadata set limit in its customization, which results in less control over the record display. Another possible improvement is related to the customized titles in the IR, Omeka, that "the clarity of human uses but the harvest machinery did not know what to do with them [the customized titles]." (Makula 204) It is central that the harvesting project ensures how metadata are housed in the ILS and how these metadata are ingested and transformed in the discovery platform for its success and the best possible discovery platform user experience. The Library seems to be one of its first ExLibris clients to purse OAI-PMH automated harvests and has a role to pave the way forward.

References

Corrado , Edward M. (2018) Discovery products and the Open Archives Initiative Protocol for Metadata Harvesting, International Information & Library Review, 50:1, 47-53.

Hardesty , Juliet L. (2014) Exhibiting library collections online: Omeka in context, New Library World, Vol. 115 Iss: 3/4, pp. 75-86.

Makula , Amanda Y. (2019) Come together: Interdepartmental Collaboration to Connect the IR and Library Catalog, The Serials Librarian, 76:1-4, 201-207.

[2] NISO ResourceSync accessed December 10, 2021 from http://www.niso.org/standards-committees/resourcesync
[3] ANSI/NISO Z39.99-2017 ResourceSync Framework Specification accessed December 10, 2021 from http://www.niso.org/publications/z3999-2017-resourcesync

Sanderson, R., Young, J., & LeVan, R. (2005) SRW/U with OAI: expected and unexpected synergies. D-Lib Magazine, 11 (2), p.1.

Omeka Journal of the Medical Library Association : JMLA Journal of the Medical Library Association : JMLA. , 2016, Vol.104(4), p.374-376

Working with Plugins
https://omeka.org/classic/docs/Admin/Adding_and_Managing_Plugins/

Omeka Plugins
https://daniel-km.github.io/UpgradeToOmekaS/omeka_plugins.html

Appendix

OAI-PMH, OAI Protocol in Remote Digital Repositories in Alma

In Ex Libri's user's manual, the features of the Remote Digital Repositories[4] are described as follows.

> Institutions can integrate their local Digital Asset Management System (such as Rosetta or DSpace, [Omeka, Fedora, Bepress/Digital Commons]) within Alma. The remote digital system must provide a bibliographic metadata record (full or partial) with the digital-specific record ID and any additional digital-oriented inventory information embedded as bibliographic field information. This feed ensures that the digital inventory from the remote system is reflected in Alma – ensuring that inventories of all types (physical, electronic, and digital) can be cataloged in one place and managed by one staff.

- The remote repositories flow manages one-way synchronization, sending metadata from the local Digital Asset Management (DAM) System to Alma.
- Alma currently supports imports in OAI formats only, either directly online through the http OAI-PMH protocol or using downloaded files in the following OAI formats: oai_dc (simple or qualified, oai_marc21 (see sample), or DSpace's simple archive format. If you are working with a different format (such as non-OAI Dublin Core, non-OAI MARC 21, or delimited text), convert it to one of the supported OAI formats before attempting to import.

Sample: oai_dc (simple or qualified)

<ListRecords>

<record>

<header>

<identifier>oai:uni/112233445561</identifier>

<datestamp>2016-01-01T09:02:44Z</datestamp>

<setSpec>Digital</setSpec>

</header>

<metadata>

[4] Ex Libris's Remote Digital Repositories https://knowledge.exlibrisgroup.com/Alma/Product_Documentation/010Alma_Online_Help_(English)/Digital_Resource_Management/020Configuring_Digital_Resource_Management/010Managing_Remote_Digital_Repositories#:~:text=Out%20of%20the%20box%2C%20Alma,DigitalCommons%20with%20selected%20OAI%20formats.

<oai_dc:dc xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/"
xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/
http://www.openarchives.org/OAI/2.0/oai_dc.xsd">

<dc:title>Title of digital record 1</dc:title>

<dc:identifier>http://www.uni.com/112233445561.ext</dc:identifier>

<dc:description>Note about file 1</dc:description>

</oai_dc:dc>

</metadata>

</record>

<record>

<header>

<identifier>oai:uni/112233445562</identifier>

<datestamp>2016-01-01T09:02:44Z</datestamp>

<setSpec>Digital</setSpec>

</header>

<metadata>

<oai_dc:dc xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/"
xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/
http://www.openarchives.org/OAI/2.0/oai_dc.xsd">

<dc:title>Title of digital record 2</dc:title>

<dc:identifier>http://www.uni.com/112233445562.ext</dc:identifier>

<dc:description>Note about file 2</dc:description>

</oai_dc:dc>

</metadata>

</record>

</ListRecords>