

5-2017

From digital repositories to the library catalogue: Two workflows for transforming metadata

Amy Leigh Allen

University of Arkansas, Fayetteville, ala005@uark.edu

Deborah E. Kulczak

University of Arkansas, Fayetteville, dkulczak@uark.edu

Mary A. Gilbertson

University of Arkansas, Fayetteville, mag@uark.edu

Follow this and additional works at: <https://scholarworks.uark.edu/libpub>

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

Recommended Citation

Allen, Amy Leigh; Kulczak, Deborah E.; and Gilbertson, Mary A., "From digital repositories to the library catalogue: Two workflows for transforming metadata" (2017). *University Libraries Faculty Publications and Presentations*. 29.
<https://scholarworks.uark.edu/libpub/29>

This Article is brought to you for free and open access by the University Libraries at ScholarWorks@UARK. It has been accepted for inclusion in University Libraries Faculty Publications and Presentations by an authorized administrator of ScholarWorks@UARK. For more information, please contact scholar@uark.edu, ccmiddle@uark.edu.

From digital repositories to the library catalogue: Two workflows for transforming metadata

Received (in revised form): 25th May, 2017



Amy Leigh Allen

is University Archivist, University of Arkansas Libraries. As the first University Archivist for the institution, she established policies and procedures for the new unit. She is responsible for collecting and processing university-related manuscript collections, providing reference service, contributing to digital exhibits and managing digital collections.

University of Arkansas Libraries, 365 N. Mclroy Ave., Fayetteville, AR 72701, USA
Tel: +1 479 575 6370;
E-mail: ala005@uark.edu



Deborah E. Kulczak

is Head of Technical Services, University of Arkansas Libraries. Prior to that, she served as the Music Cataloger, Database Maintenance Librarian and Non-MARC Metadata Coordinator for the Libraries, in which position she was responsible for creating descriptive metadata and assigning controlled vocabulary for digital exhibits.

University of Arkansas Libraries, 365 N. Mclroy Ave., Fayetteville, AR 72701, USA
Tel: +1 479 575 4811;
E-mail: dkulczak@uark.edu



Mary A. Gilbertson

is the Monographs Cataloging and Acquisitions Unit Head at the University of Arkansas. She is responsible for the purchasing and cataloguing of books. She has also been involved in setting up the workflow of many of the record-loading processes.

University of Arkansas Libraries, University of Arkansas Libraries, 365 N. Mclroy Ave., Fayetteville, AR 72701, USA
Tel: +1 479 575 5417;
E-mail: mag@uark.edu

Abstract Making a connection between digital repositories and library catalogues is an issue many libraries and archives face, especially when dealing with items that traditionally received full cataloguing records. This paper explores different workflows for reusing metadata to create catalogue records, connecting the two resources together. In order to have one interface to search both the older physical items and newer digital items, the archives and technical services department of the University of Arkansas Libraries worked together to create different workflows to save time and eliminate the need for double entry, using the open source software MarcEdit and XML Notepad. Two different processes were created for different types of materials in different repositories. The first process transforms metadata for undergraduate theses in DSpace from Dublin Core to MARC.

The second process transforms metadata from music concert recordings in Islandora from MODS to MARC. Readers will learn multiple options and techniques for converting metadata to MARC records.

KEYWORDS: digital asset management, metadata, MARC, theses, dissertations, concert recordings, institutional repository

INTRODUCTION

Making a connection between digital repositories and library catalogues is an issue many libraries and archives face, especially when dealing with items that have traditionally received full cataloguing records. On the one hand, there is the desire to provide a single interface for searching both the older physical items and newer digital items. On the other hand is the desire to eliminate the double metadata entry that can be required to achieve that goal.

This paper explores two different workflows developed at the University of Arkansas Libraries for reusing repository metadata to create catalogue records, connecting the two resources together. The first process transforms metadata for undergraduate theses in a DSpace repository from Dublin Core (DC) to MARC. The second process transforms metadata from music concert recordings in an Islandora repository from Metadata Object Description Schema (MODS) to MARC. These processes take advantage of open source software programs such as MarcEdit, XML Notepad and Google Documents, although the proprietary Oxygen XML editor is very helpful for editing and testing XSLT stylesheets.

The paper demonstrates the benefits to be gained when archivists, cataloguers and other librarians collaborate. For, as Terry Reese, the developer of MarcEdit, said in a 2009 article:

as more institutions bring digital collections online, technical services staff will continue

to face the growing issue of distributed metadata retrieval. Unlike their print cousins, today's institutional repositories and digital collections give rise to metadata of a distributed nature that require technical services departments to think creatively and produce workflows that encourage repurposing data.¹

It is hoped that the work described following can serve as but one example.

SELECTED LITERATURE REVIEW

There are various articles in the library literature that discuss an automatic or semi-automatic conversion of author-supplied metadata to a MARC record, a few of which are summarised here. Ken Robinson, Jeff Edmunds, and Stephen C. Mattes point out that most academic theses and dissertations are now born-digital assets.² As such, they often coexist with author-supplied metadata with the potential to be repurposed. The authors provide explanations of the harvesting process, metadata transformation and record loading using MarcEdit's Metadata Harvester.

Two articles discuss using a Perl script for this process. Brian E. Surratt and Dustin Hill discuss using a Perl script for cataloguing theses and dissertations at Texas A&M.³ In this process, the Perl module MARC::Record is used to turn digital metadata into MARC records. This Perl script was written as a collaboration between the library's electronic resources cataloguer and the Thesis Office's programmer analyst,

who performs ongoing maintenance. The records are edited by cataloguers to conform to Anglo-American Cataloguing Rules (AACR2) standards and then saved to WorldCat with OCLC encoding level 'K' (minimal level records). In 2008, Sevim McCutcheon *et al.* also discuss using a Perl script for cataloguing theses and dissertations at Kent State.⁴ The Perl program uses the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) to extract metadata, modify and enhance the data, and insert the data into their Innovative Interfaces catalogue. They retrieve metadata using Library for WWW in Perl (LWP). The records go through final editing in the OCLC Connexion client (OCLC's cataloguing interface) and are sent to OCLC WorldCat.

In an article from 2009, Michael Boock and Sue Kunda⁵ discuss cataloguing theses and dissertations at Oregon State University. This paper states that MarcEdit is used to transform DC to MARC, but the details are not revealed. This paper focuses on the comparison between workflows for print theses and dissertations vs electronic theses and dissertations (ETDs). After the metadata are converted, library staff add Library of Congress subject headings (LCSH) to each record.

In the aforementioned article from 2009,⁶ Reese, also of Oregon State, examines the challenges posed by the growing number of born-digital resources for technical services departments. Fortunately, for those seeking to describe these resources with MARC records, there are several tools available — including Reese's MarcEdit, Innovative's XML Harvester and OCLC Connexion — that can be used for harvesting and transforming repository metadata for use in library catalogues. Reese includes his complete XSLT stylesheet for converting OAI records to MARC XML (the schema for representing MARC data in XML).

Also in 2009, Shawn Averkamp and Joanna Lee⁷ describe the University of Iowa Libraries' use of an XSLT stylesheet to transform ProQuest XML metadata for ingest into their Digital Commons (bepress) institutional repository. They also discuss their procedures for using MarcEdit and its built-in crosswalks to generate MARCXML, and subsequently MARC records, from bepress for their online library catalogue. (The field mapping and XSLT stylesheets for both segments of the process are provided with the article.)

In an article from 2016, Marielle Veve⁸ writes about some of the most popular automated and semi-automated approaches currently used in libraries to harvest electronic theses and dissertations. OCLC Digital Gateway is mentioned as the only fully automated tool, but Veve states that with Digital Commons (bepress) or DSpace 'the OCLC Digital Gateway proved it could not harvest Qualified DC feeds as claimed, but only Unqualified DC'. Veve mentions semi-automated approaches that rely on ProQuest services, semi-automated approaches that rely on highly technical tools, and semi-automated approaches that rely on the MarcEdit OAI Harvester. Veve's conclusion is that the methods she outlines will all need customisations for each library.

In a related 2016 article, Veve⁹ describes the methods used at the University of North Florida for creating and harvesting ETD metadata from their Digital Commons (bepress) repository and repurposing it in OCLC — notably, without making use of ProQuest metadata. (UNF's field mapping and XSLT stylesheets are also provided.)

While there are a number of articles devoted to crosswalking and repurposing metadata for electronic theses and dissertations, little can be found on the topic as related to streaming audio files. Riley, Mullin and Hunter¹⁰ describe a

process under development at Indiana University in 2009 for batch converting MARC bibliographic and authority records for use in their Variations system. (Variations was the software developed for the university's digital music library.) As will be seen, however, the University of Arkansas Libraries was looking to take its metadata in the opposite direction, starting with MODS in its repository and converting to MARC for the online catalogue. The Library of Congress does post a MODS to MARC conversion map on its website,¹¹ but the authors could find no articles describing implementations for audio files. In retrospect, this is perhaps not surprising, given that MODS was originally developed by the Library of Congress

'to satisfy the expressed need for an abbreviated XML version of MARC 21 [the current version of MARC] . . . [and was] intended to carry selected data from existing MARC 21 records as well as to enable the creation of original resource description records'.¹²

UNDERGRADUATE THESES

Background

In 2010, the University Archivist was approached to partner on the management of digital objects that had been previously collected. The Engineering and Mathematics Librarian had been collecting undergraduate honours theses from the Engineering College for roughly five years previous. While there were already established practices for the Libraries to collect and catalogue graduate theses and dissertations, undergraduate theses had historically been maintained by each individual college. The Engineering College, however, was transitioning to accept theses in digital format only and expressed a desire for the libraries to collect and provide access to the undergraduate theses. The head of the Libraries' Systems Department also

approached the University Archivist with the desire to set-up and test a DSpace repository. The platform was a natural fit for the undergraduate theses and a pilot project was begun.

Librarians and archivists wanted the information used for the metadata of the undergraduate theses to correspond with the information used in the MARC fields for graduate theses. A working group was formed within the Libraries with the University Archivist and two cataloguers from the Technical Services Department. The working group used catalogue records from graduate theses as a template to define the metadata fields for the undergraduate theses. As DSpace uses DC metadata, a guide was created to designate which DC fields would be used, along with a definition and example of the field. The guide also designated how the DC fields would be mapped to MARC.

The guide was altered within the first year when the Libraries in coordination with University Information Technology Services decided to adopt Vireo software as an electronic submission system. Vireo is a software developed by the Texas Digital Library in collaboration with Texas A&M University for the ingest of electronic theses and dissertations (<https://tdl.org/etds/>). Vireo uses the metadata schema defined by the Networked Digital Library of Theses and Dissertations (NDLTD). This schema is made up of mainly DC elements with a few elements specific to theses and dissertations not found in DC, including <thesis.degree.level>, <thesis.degree.name>, <thesis.degree.department>, <thesis.degree.discipline> and <thesis.degree.grantor>. The non-DC fields were also mapped to MARC (see Appendix A).

Once the test was successful with the backlog of engineering undergraduate theses, the programme was opened up to students from all five colleges across campus. Beginning in December 2011, current students could submit an undergraduate

thesis directly to the Libraries through an online system.

First workflow

The theses had to be dealt with in two groups: the backlog of digital files and the new theses which would be uploaded through Vireo. The backlog digital files were loaded directly into DSpace with title as the only metadata. The titles and digital object identifiers (DOIs) were exported to a spreadsheet.

Fields were then added to the spreadsheet for the rest of the desired metadata.

A cataloguer in the Technical Services Department created the rest of the metadata based on information that could be gleaned from the thesis itself. The cataloguer sent the spreadsheet back to the University Archivist with the completed metadata, which was then loaded from the spreadsheet into DSpace to correspond with the previously loaded PDFs.

Once the electronic submission system was instituted in 2011, information for the metadata was submitted by the student along with the thesis file. Students navigated to an online interface using Vireo software, filling in a series of informational fields, including title, graduation date, adviser, degree name, etc., which would become the metadata associated with the thesis PDF file. The thesis was reviewed and approved by an administrator, usually within the college associated with the student, and the thesis was then transferred from Vireo to DSpace. At the end of each semester, the University Archivist exported the metadata from DSpace as a CSV file. The spreadsheet was then edited using Google Document software in order to maintain the UTF-8 encoding for diacritics. The DSpace specific fields not needed for cataloguing were deleted, including <dc.date.accessioned>, <dc.date.available> and <dc.date.updated>. The University Archivist then sent the

spreadsheet to a cataloguer. Figure 1 provides an example of a record from DSpace in DC created by this process.

The process for converting the DC to MARC was the same for both the backlog and files going forward. The cataloguer re-saved the file as an XLSX file but still used Google Documents open software to format and further edit the metadata: deleting additional columns, adding columns for further information and verifying all information was in the proper format for the catalogue record. The cataloguer next used the Mail Merge feature in Microsoft Word to export the data into a Word template which added MARC identifiers. Once in Word, the author field was edited to fit AACR2 formatting. Additional fields were added for department or subject headings (if needed), and blank fields were deleted. The cataloguer verified that there was a line space (hard return) at the end of each record, and the file was saved as a TXT file with UTF-8 encoding. The cataloguer used the MarcMaker function of the MarcEdit software to convert the TXT file to MARC with MARC-8 encoding. The MARC records were then loaded into the Libraries' Innovative Interfaces (III) catalogue. Approximately 183 undergraduate honours theses were transformed over a two year period using this process. Figure 2 provides an example of a MARC record created by the first workflow.

Innovative Interfaces systems make use of load tables for the batch manipulation and processing of incoming MARC data. These load tables can be created or edited locally by the two University Libraries cataloguers who have completed the vendor's required training. This gives the cataloguers a great deal of flexibility, in that they can determine if it is more expedient for a given data source to create a customised load table, to use MarcEdit alone, or to employ a combination. In the case of the honours theses, no specialised load table was required.

Figure 1 Honours thesis, DSpace DC output.

dc.contributor.advisor	Haggard, Brian E.
dc.creator	Bailey, Iain M.
dc.date.accessioned	2012-04-13T18:42:28Z
dc.date.available	2012-04-13T18:42:28Z
dc.date.created	2012-05
dc.date.issued	2012-04-13
dc.date.submitted	May 2012
dc.identifier.uri	http://hdl.handle.net/10826/ETD-2012-05-74
dc.description.abstract	<p>The addition of chemical amendments to land applied poultry litter is paramount to the maintenance of acceptable phosphorus (P) and nitrogen (N) levels stormwater runoff from pasture. Chitin is the chief component of shrimp shells, and together with its more processed form chitosan, may show promising P, N, and metal ion chelation characteristics by virtue of its chemical structure and unique charge distribution. The objective of this study was to measure the effects of aluminum sulfate (alum, Al₂(SO₄)₃) on aqueous solubility of P and ammonia (NH₃) release against those of chitin and three varieties of chitosan in poultry litter.</p> <p>The experiments revealed that while chitin has little to no effect on water extractable P (WEP), all three varieties of chitosan show decreases in WEP not significantly different from those of alum. While alum significantly decreased NH₃ volatilization from litter samples, neither chitin nor chitosan significantly affected released NH₃.</p>
dc.format.mimetype	application/pdf
dc.language.iso	eng
dc.subject	Agriculture
dc.subject	Environmental engineering
dc.subject	Agriculture engineering
dc.title	Chitosan reduces water solubility of phosphorus in poultry litter
dc.date.updated	2012-04-13T18:42:34Z
dc.identifier.slug	10826/ETD-2012-05-74
thesis.degree.name	Bachelor of Science in Biological Engineering
thesis.degree.level	Undergraduate

thesis.degree.grantor	University of Arkansas, Fayetteville
thesis.degree.discipline	Biological Engineering
thesis.degree.department	Biological and Agricultural Engineering
dc.type.material	text
dc.type.genre	thesis
dc.contributor.committeeMember	Liang, Yi
dc.contributor.committeeMember	Zaharoff, David A.

Figure 2 Honours thesis, MARC record, first workflow.

```

LEADER 00000nam a22000003a 4500
005 20120619153015.0
006 m d
007 cr un ---uuuuu
008 061912s2012 xx o 000 0 eng d
049 AFUT
100 1 Bailey, Iain M
245 10 Chitosan reduces water solubility of phosphorus in poultry
litter|h[electronic resource] /|cby Iain M. Bailey
260 |c2012
300 1 online resource
500 "May 2012"
502 Thesis (Honors)--University of Arkansas, Fayetteville, 2012
520 3 The addition of chemical amendments to land applied poultry
litter is paramount to the maintenance of acceptable phosphorus
(P) and nitrogen (N) levels in stormwater runoff from pasture.
Chitin is the chief component of shrimp shells, and together
with its more processed form chitosan, may show promising
P, N, and metal ion chelation characteristics by virtue
of its chemical structure and unique charge distribution.
The objective of this study was to measure the effects of
aluminum sulfate (alum, Al2(SO4)3) on aqueous solubility of P
and ammonia (NH3) release against those of chitin and three
varieties of chitosan in poultry litter. The experiments
revealed that while chitin has little to no effect on water
extractable P (WEP), all three varieties of chitosan show
decreases in WEP not significantly different from those of
alum. While alum significantly decreased NH3 volatilization
from litter samples, neither chitin nor chitosan significantly
affected released NH3
538 System Requirements: Internet access, World Wide Web browser,
Adobe Acrobat Reader. Mode of access: World Wide Web
655 7 Electronic theses.|2AFU
690 Dept.: Biological and Agricultural Engineering
690 Advisor: Haggard, Brian E
690 Agriculture
690 Environmental engineering
690 Agriculture engineering
830 0 UA undergraduate honors paper
856 40 |uhttp://hdl.handle.net/10826/ETD-2012-05-74

```


Figure 3 Honours thesis, MarcEdit conversion, second workflow.

```

=LDR 01156njm a2000145a 4500
=042 \\$adc
=720 \\$aFleming, Robert$eauthor
=690 \\$aZou, Min$econtributor
=690 \\$aMechanical Engineering
=260 \\$c2009-05
=520 \\$aOne- and two-dimensional silicon nanostructures, such as
      nanodots,nanowires, and surface topography, have potential
      uses in a variety of fields, including microelectronics,
      photonics, and tribology. In this thesis research,
      nickel-induced crystallization (NIC) of amorphous silicon
      (a-Si) was investigated to determine the suitability of NIC
      of a-Si to produce silicon nanostructures on silicon substrates.
      The results show that formation of both surface topography
      and nanowires is possible with NIC of a-Si. In addition, the
      formation of nanostructures is very sensitive to Ni thickness,
      and the resulting nanostructures and films display a high degree
      of repeatability. The results obtained from this research
      provide a solid foundation for future studies in NIC of a-Si.
=024 8\\$ahhttp://hdl.handle.net/10826/71
=653 \\$aMechanical engineering
=245 00$aNickel-induced crystallization of amorphous silicon

```

Second workflow

The working group sought to simplify and improve the process with the intent of replacing the first workflow. The group was inspired by Deng and Reese's 2009 article¹³ on mapping metadata from DSpace to OCLC and began testing methods for harvesting the metadata directly from DSpace into MARC. While many metadata harvesters are available, MarcEdit was chosen as the best choice as current staff are already familiar with the software and use it regularly to modify bibliographic records within record sets in the online catalogue. MarcEdit is also freely available and the programmer, Terry Reese, provides documentation and support for the software.

The University Archivist and cataloguer tested the harvest with the stylesheet that comes standard in MarcEdit. While the standard DC fields were successfully harvested, a different option was needed for the fields specific to the Networked Digital Library of Theses and Dissertations.

This would require editing the XSLT stylesheet. A student worker offered to assist with editing the XSLT stylesheet. Figure 3 provides an example of the data harvested using MarcEdit.

Before the customised XSLT stylesheet was completed, the campus decided to adopt Digital Commons (bepress) as the campus-wide institutional repository, and it was determined that the undergraduate theses would be moved from DSpace to Bepress, and the metadata and harvesting options were put on hold until this transition was completed. The working group will continue to work on metadata harvesting options.

CONCERT RECORDINGS

Background

The Music Department at the University of Arkansas has been recording concerts performed by its students and faculty since the 1960s. The department has consistently

donated copies of these concerts to the University Libraries since the 1980s. The most recent physical format of these concerts was on compact disc (CD). In 2012, the Music Department announced steps to move to an all-digital process. The department would no longer produce CDs and instead would donate copies of recordings as digital files, which the Music Department wanted streamed online. In the last few years, this has amounted to slightly over 100 titles annually.

To address this change, a working group was formed between the University Archivist, the then Music Cataloguing Unit Head, and the then Head of the Performing Arts and Media Library. At this time, the Libraries had access to two different repository platforms: DSpace and Islandora. Islandora is a Fedora platform with a Drupal interface (<https://islandora.ca/>). A test music concert was added to each platform to assess features. It was determined that Islandora offered the greatest number of desired features: the ability to automatically create MP3 files from WAV files, the ability to stream the recordings within the Islandora interface, and an easy batch import. The Islandora platform offered two metadata schema choices: simple DC or MODS, and it was agreed by the working group that the more precise MODS schema would be better suited for this project.

In 2003, Rebecca Guenther and Sally McCallum of the Library of Congress characterised the emerging schema with these words:

the need for a rich metadata standard such as MODS has been expressed by members of the digital library and related communities as they attempt to implement projects involving search and retrieval, management of complex digital objects, integrating metadata from library databases with other non-MARC sources and other functions.¹⁴

Several authors — including Edge and Han,¹⁵ Herring,¹⁶ Hoebelheinrich¹⁷ and Park and Maszaros¹⁸ — have subsequently affirmed the suitability of MODS as the descriptive metadata standard for non-textual, and particularly audio, resources. Factors they have cited echo Guenther and McCallum and include: the intuitiveness of MODS for library professionals equipped with even basic cataloguing knowledge; a rich set of semantics that can be used in combination with other metadata standards, including METS; its capabilities for recording provenance information and multiple access points; its ability to convey complex relationships between resources; and the flexibility to customise it locally, either through the <extension> element or ‘type’ attributes. MODS is especially useful for describing musical compound objects, in that it can preserve the hierarchical relationship between the recording as a whole and its component tracks. This can be achieved using the <relatedItem> element for each individual track, within which any of the MODS elements and sub-elements (<titleInfo>, <name>, <originInfo>, etc.) can be repeated.

Once the selection of MODS was made at the University of Arkansas, again librarians and archivist worked together to define specifically which MODS fields would be used. They also provided a definition and example of each field, along with how each field would be mapped to MARC for cataloguing (see Appendix B).

The Libraries had been creating MARC records, albeit brief ones, for concert recordings since 1999. At the time of the switch to streaming audio, more than 2,600 recordings dating back to 1963 had been added to the library catalogue. (Local practice has been not to add the records to OCLC WorldCat.) The original MARC record template and instructions for creating the concert metadata were developed by the head of Performing Arts and Media (PAM) and the music cataloguer. In some ways,

these brief records conformed to standard music cataloguing practice, in that they contained contents notes, recording date, and added entries for the performers (including role terms). Because the intent was to have paraprofessional staff in PAM create the records on an ongoing basis, notably absent were the name/uniform title access points for the works contained in the recording. Subject headings were assigned from a locally created list (eg 'Voice recital', 'Graduate recital') and entered in MARC field 690.

Where these brief records differed most from standard practice was in the title fields. The standard phrase '[Concert recording]' was chosen for the MARC 245 field to allow staff and patrons alike to easily pull up a list with a catalogue search. Alternative titles (eg 'Wednesday recital hour', 'Black music symposium') were recorded in MARC 246 fields when deemed important for access. Another divergence was the local decision to record the Music Department as the creator (MARC 110 field) in every case. Finally, because patrons most often requested concert recordings by performance date, an access point for this was provided in the form of a local call number (099 field). The date was preceded by a designation for the format (CCD for compact disc, CCT for cassette tape), and when multiple concerts were

recorded on a single day, work letters such as 'a', 'b', etc. were added to the end of the date. (The practice of using the CCD prefix was continued for streaming recordings, because occasionally the Libraries have also received an accompanying CD.) See Figure 4 for a brief manually keyed MARC record.

Librarians and archivists wanted the new digital concert recordings to continue to be catalogued so that both old and new formats could be searched in a single interface. What they did *not* want was to make additional work by creating both MODS records and MARC records from scratch to import into two different systems. The working group therefore set about developing a workflow where one record could be created and used in both systems.

Workflow

As was previously mentioned, the catalogue records for physical CDs were being created by a staff member in the Performing Arts and Media Library. Supervisors agreed that this same staff member would be in charge of processing the digital files in the new workflow, including creating metadata and uploading the files to Islandora, which would replace physical processing and filing the CDs.

Figure 4 Concert recording, brief record keyed by staff.

```
LEADER 00000njm 2200000 a 4500
007 ss
099 CCT 861029
110 University of Arkansas, Fayetteville. Dept. of Music
245 0 [Concert recording]
246 Wednesday recital hour
260 |c1986
505 Suite gothique / L. Boellmann -- Dixhuitieme ordre / Francis
Couperin
518 Recorded Oct. 29, 1986
690 Organ recital
690 Harpsichord recital
700 Adams, James, organ
700 Clinkscapes, Joyce, harpsichord
```

An important piece of information was needed from Islandora: the digital object identifier (DOI), which was automatically generated through a handle server linked to Islandora. The DOI would be used as the location/URL link in the catalogue record for the user to listen to the recording. Therefore, the MODS metadata record was created first. It was necessary to use the batch import feature, as the Music Department was recording each song track as a separate digital file, so that one concert would be made up of a series of digital files. In addition to the music files, a PDF of the concert programme was submitted along with the audio recordings.

All of the audio files plus the programme for each concert were then grouped together into one folder or collection in Islandora. The metadata could subsequently be reused for the different files associated with a single concert with just the title changed for the individual files. What this meant in actual practice was that for the MODS metadata, it was decided that '[Concert recording]' would be coded in the <titleInfo><subTitle> element for both the object as a whole and for each individual file. The <titleInfo><title> element would be reserved for the concert date (eg '2015-04-28') in the object level metadata and 'Track 01', 'Program', etc. in the metadata for the individual files. The rest of the object-level metadata would simply be repeated in the individual descriptions. Note that the Libraries' approach did not take advantage of MODS' capability for establishing hierarchical structure using the <relatedItem> element. Rather, Islandora itself provides this structure. It also generates DC metadata on the fly for use in OAI-PMH harvesting.

The working group created a template for the metadata in MODS XML. What follows is the current procedure for generating the MODS and MARC records. The staff member in the Performing Arts and Media Library creates the metadata for the individual concerts using the template

and the free software XML Notepad. The MODS XML metadata files, music files, and the PDF of the concert programme are then all packaged into a zip file which can be batch uploaded into Islandora. During the uploading process, the DOI is generated, and the staff member adds this information to the MODS XML metadata file. See Figure 5 for an example of a record from Islandora in MODS created by this process.

Next, the MODS XML files are converted to MARC files using another free software program — MarcEdit. For each recording, the MODS object-level description is output from Islandora. A batch of MODS XML files are grouped together for processing into one folder on the computer, and the MarcEdit 'Batch Process Records' function is used to convert them simultaneously from MODS XML to MARC. Originally, cataloguers used the MODS=>MARC XSLT stylesheet (file name 'MODS2MARC21slim.xml') supplied with the software and made just a single edit — to allow the <location><shelfLocation> element to come in as the local call number field. The rest of the desired metadata was inserted or transformed using a custom III load table. Cataloguers in the Technical Services Department, however, have been seeking to expand their knowledge of XSLT, and recently, with the help of the Libraries' web developer, the MarcEdit stylesheet was refined and can complete the transformation independent of any load table. (This stylesheet is available as Appendix C). For those interested in using this stylesheet, it should be noted that the III system automatically sorts fields of incoming records in MARC tag order; therefore, developers did not attempt to arrange the stylesheet in MARC tag order. Furthermore, concert recording MARC records have not yet been upgraded to conform to Resource Description & Access (RDA) standards. The cataloguers hope to make this change in the near future. See Figure 6 for an example of the bibliographic output.

Figure 5 Concert recording, Islandora MODS output.

```
<mods xmlns="http://www.loc.gov/mods/v3" xmlns:mods="http://
www.loc.gov/mods/v3" xmlns:xlink="http://www.w3.org/1999/xlink"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
  <titleInfo>
    <title>2014-01-31</title>
    <subTitle>[Concert recording]</subTitle>
  </titleInfo>
  <titleInfo type="alternative">
    <title>Black Music Symposium Second Annual Concert
    </title>
  </titleInfo>
  <name type="corporate">
    <namePart>University of Arkansas, Fayetteville</namePart>
    <namePart>Dept. of Music</namePart>
  </name>
  <name type="corporate">
    <namePart>University of Arkansas, Fayetteville</namePart>
    <namePart>Schola Cantorum</namePart>
  </name>
  <name type="personal">
    <namePart>Caldwell, Stephen</namePart>
    <role>
      <roleTerm type="text">conductor</roleTerm>
    </role>
  </name>
  <name type="corporate">
    <namePart>Rick Salonen Ensemble</namePart>
  </name>
  <name type="personal">
    <namePart>Kahng, Er-Gene</namePart>
    <role>
      <roleTerm type="text">violin</roleTerm>
    </role>
  </name>
  <name type="personal">
    <namePart>Botes, Johan</namePart>
    <role>
      <roleTerm type="text">piano</roleTerm>
    </role>
  </name>
  <name type="personal">
    <namePart>Park, Moon Sook</namePart>
    <role>
      <roleTerm type="text">soprano</roleTerm>
    </role>
  </name>
</mods>
```

```

<name type="personal">
  <namePart>Mains, Ronda</namePart>
  <role>
    <roleTerm type="text">flute</roleTerm>
  </role>

```

Figure 6 Concert recording MarcEdit conversion.

```

=LDR 01907njm a2200337uu 4500
=006 m\\o\\h\\
=007 cr|
=007 sz|
=008 \\s2014\\xxu|o|
=049 \\$aAFUT
=500 \\$aStreaming audio.
=245 10$a[Concert recording].
=246 3$aBlack Music Symposium Second Annual Concert.
=110 2$aUniversity of Arkansas, Fayetteville.$bDept. of Music.
=710 2$aUniversity of Arkansas, Fayetteville.$bSchola Cantorum.
=700 1$aCaldwell, Stephen,$econductor.
=710 2$aRick Salonen Ensemble.
=700 1$aKahng, Er-Gene,$eviolin.
=700 1$aBotes, Johan,$epiano.
=700 1$aPark, Moon Sook,$esoprano.
=700 1$aMains, Ronda,$eflute.
=700 1$aDelaplain, Theresa,$eoboe.
=700 1$aKashiwagi, Tomoko,$epiano.
=505 0$a[Track 01]. The voice of the lord, Psalm 29 / William Grant
Still -- [Track 02]. The mooche & East St. Louis Toodle-oo
/ Edward K. "Duke" Ellington -- [Track 03]. Suite for violin
and piano. Suggested by Richmond Barthe, "African Dancer" ;
[Track 04]. Suggested by Sargent Johnson, "Mother and child" ;
[Track 05]. Suggested by Augusta Savage, "Garmin D multiple"
/ William Grant Still -- [Track 06]. Selections. Songs to the
dark virgin ; [Track 07]. Night ; [Track 08]. My soul's been
anchored in the lord / Florence Price -- [Track 09]. Miniatures.
I ride an old paint ; [Track 10]. Adolorido ; [Track 11]. Jesus
is a rock in the weary land ; [Track 12]. Yaravi ; Froggy went
a courtin' / William Grant Still -- [Track 13]. Elijah rock /
arr. Moses Hogan.
=518 \\$aRecorded January 31, 2014.
=260 \\$c2014.
=500 \\$aTrack 12 contains both Yaravi and Froggy went a courtin' from
Miniatures by William Grant Still.
=690 \\$aVocal music recital.
=690 \\$aInstrumental music recital.
=856 40$uhttp://hdl.handle.net/10975/islandora:193427
=099 \\$aCCD 140131

```

Before loading the new MARC files, one final MarcEdit function is used to make the process easier — ‘MARCJoin’, which combines the still individual MARC records into a single file. Loading into the III system can then proceed, and once finished, the concert recordings are searchable in both the Libraries’ catalogue and through the Islandora interface.

CONCLUSION

These processes represent the initial attempts at the University of Arkansas Libraries to transform metadata from local digital repositories into catalogue records using a semi-automated process. As stated, there are significant benefits to be realised when cataloguers, archivists and other librarians collaborate on metadata projects. But what does this mean specifically for the University of Arkansas Libraries experience? First, such collaboration reinforced the idea that the knowledge and perspectives of those who practise different specialties and serve varying clientele make for a stronger end product. Furthermore, when there is collaboration from the onset — with the selection of metadata schemas, evaluation of desired metadata elements, and even the choice of the repository — there are fewer missteps along the way. Finally, the authors gained a renewed appreciation for the collaborative nature of the library profession as a whole, as represented by developers like Terry Reese who offer their tools as open source software and who work with the user community to refine them constantly, or by the librarians and archivists who share their own knowledge and expertise in the published literature, enabling their colleagues anywhere in the world to adapt and use what they have produced. Through these various modes of collaboration, it was possible to streamline the process of metadata creation for digital undergraduate theses and Music Department concert recordings for both the native repository and the Libraries’ catalogue.

But the importance of collaboration is just one takeaway from this experience. Another is that change is a constant, and change will usually require rethinking metadata workflows. One (discussed earlier) was the campus decision to implement Digital Commons (bepress) as the repository for undergraduate theses, rather than continuing to use DSpace. But there were two other changes that also had a direct bearing on this metadata work. Due to a pending evaluation of the sustainability of open source software on campus, uploading and streaming concert recordings using Islandora is now on hold, and the Libraries have temporarily reverted to making physical CDs and entering metadata directly into the catalogue.

Additionally, during the past year, the Libraries implemented a discovery layer on top of the existing catalogue. This discovery software can crawl institutional repositories, bring repository and catalogue metadata together, and serve them in a single ‘QuickSearch’. Given this latest development, one could argue that some of the processes developed at the University of Arkansas are now moot. The authors, however, do not feel this to be the case. To date, only ScholarWorks (bepress) and CONTENTdm collections have been brought into the fold, omitting Islandora. Moreover, the music reference staff will continue to direct their patrons to the library catalogue for the foreseeable future because it is more precise for locating performed and notated music. Leaving out the concert recordings would be a disservice to those users. At the very least, although, there should be an evaluation of what impact the discovery layer might have on these metadata processes.

A final — and very valuable — lesson is that acquiring new skills is never a waste of time in metadata work. Having success with these projects provided a base to allow the Libraries’ faculty and staff to keep adapting and building additional processes as our repositories change and grow. The University Archivist, cataloguers and

institutional repository staff are currently working to redesign the methods used for honours theses to the graduate theses and dissertations (approximately 400 titles submitted per year). Employing some of very XSLT stylesheets found in the literature cited for this paper, they have arrived at a process to use the metadata generated by ProQuest when students upload their work. This metadata contains nearly all the elements needed to populate ScholarWorks, and hidden fields can be brought along for repurposing in the MARC records. MarcEdit can then be used to harvest DC metadata from ScholarWorks, and a second customised XSLT stylesheet will transform it to MARC. Aside from basic quality control, the group anticipates having only to change the creator headings to RDA form and assign Library of Congress classification numbers and subject headings. (The subject librarians were recently consulted about whether this access was still desirable, and the answer was a resounding ‘yes’.) Once the process is refined for the graduate theses and dissertations, this workflow will also be used for the undergraduate honours theses. The practice of developing the procedures explained in this paper has prepared those involved to adapt to the changing landscape of metadata creation.

REFERENCES

1. Reese, T. (2009) ‘Automated metadata harvesting: low-barrier MARC record generation from OAI-PMH repository stores using MarcEdit’, *Library Resources & Technical Services*, Vol. 53, No. 2, pp. 121–134.
2. Robinson, K., Edmunds, J. and Mattes, S.C. (2016) ‘Leveraging author-supplied metadata, OAI-PMH, and XSLT to catalog ETDS’, *Library Resources & Technical Services*, Vol. 60, No. 3, pp. 191–203.
3. Surrat B. and Hill D. (2004) ‘ETD2MARC: a semiautomated workflow for cataloging electronic theses and dissertations’, *Library Collections, Acquisitions, & Technical Services*, Vol. 28, No. 2, pp. 205–223.
4. McCutcheon, S., Kreyche, M., Beecher, M.M. and Nickerson, J. (2008) ‘Morphing metadata: maximizing access to electronic theses and dissertations’, *Library Hi Tech*, Vol. 26, No. 1, pp. 41–57.
5. Boock, M. and Kunda, S. (2009) ‘Electronic thesis and dissertation metadata workflow at Oregon State University Libraries’, *Cataloging & Classification Quarterly*, Vol. 47, No. 3–4, pp. 297–308.
6. Reese, ref. 1 above.
7. Averkamp, S. and Lee, J. (2009) ‘Repurposing ProQuest metadata for batch ingesting ETDS into an institutional repository’, *Code4lib Journal*, Vol. 7, No. 3, available at: <http://journal.code4lib.org/articles/1647> (accessed 8th January, 2017).
8. Veve, M. (2016) ‘Harvesting ETD metadata from institutional repositories to OCLC: approaches and barriers to implementation’, *Journal of Library Metadata*, Vol. 16, No. 2, pp. 69–79.
9. Veve, M. (2016) ‘From Digital Commons to OCLC: a tailored approach for harvesting and transforming ETD metadata into high-quality records’, *Code4lib Journal*, Vol. 33, No. 1, available at: <http://journal.code4lib.org/articles/11676> (accessed 28th December, 2016).
10. Riley, J., Mullin C. and Hunter C. (2009) ‘Automatically batch loading metadata from MARC into a work-based metadata model for music’, *Cataloging & Classification Quarterly*, Vol. 47, No. 6, pp. 519–543.
11. Library of Congress. (2015) ‘MODS 3.5 to MARC 21 mapping, version 3.4’, Library of Congress, Washington, DC, available at: <http://www.loc.gov/standards/mods/v3/mods2marc-mapping.html> (accessed 27th December, 2016).
12. Guenther R. and McCallum S. (2002) ‘New metadata standards for digital resources: MODS and METS’, *Bulletin of the American Society for Information Science & Technology*, Vol. 29, No. 2, pp. 12–15.
13. Deng, S. and Reese, T. (2009) ‘Customized mapping and metadata transfer from DSpace to OCLC to improve ETD work flow’, *New Library World*, Vol. 110, No. 5/6, pp. 249–264, available at: <https://doi.org/10.1108/03074800910954271> (accessed 8th January, 2017).
14. Guenther and McCallum, ref. 12 above.
15. Edge, R. and Han, M. (2013) ‘Building a sustainable metadata workflow for audio-visual resources: University of Illinois Library’s Medusa Digital Preservation Repository’, IFLA World Library and Information Congress, Singapore, available at: <http://library.ifla.org/94/1/124-edge-en.pdf> (accessed 29th December, 2016).
16. Herring, M. (2015) ‘An application profile of MODS to describe complex digital musical audio resources’, *Journal of Library Metadata*, Vol. 15, No. 2, pp. 63–78.
17. Hoebelheinrich, N.J. (2006) ‘Gathering audio metadata for the Monterey Jazz Festival concerts’, Online Audiovisual Catalogers Conference, Mesa, AZ, October, available at: http://www.loc.gov/standards/mets/presentations/nancysolac/OLAC_2006_MJFppt (accessed 29th December, 2016).
18. Park, J. and Maszaros, S. (2009) ‘Metadata Object Description Schema (MODS) in digital repositories: an exploratory study of metadata use and quality’, *Knowledge Organization*, Vol. 36, No. 1, pp. 46–59.

Appendix A: DC to MARC crosswalk for honours theses

DSpace			Marc 21	
field name	Dublin Core name	Note	field	Example
Abstract	description.abstract	To be written by the author, otherwise leave blank	520	Include full abstract
Authors	creator	Use name authority record	100	Cloud, Andrew N.
Coverage	coverage.spatial	Can be used if research location (other than Arkansas) was instrumental for research; Use authority record		
Date Added	date.accessioned	Automatically added		
Date Available	date.available	Automatically added		
Date Issued	date.issued	Automatically added		
Date Created	date.created	Use year the author graduated	500	May 2007
Description	description	May use if no abstract was written or for information not appropriate for an abstract, such as awards received		
Format	format	Automatically added		

DSpace			Marc 21	
field name	Dublin Core name	Note	field	Example
Identifiers	identifier.uri	Automatically added	856	http://hdl.handle.net/ 10826/ETD-2009-08-44
Language	language	Choose language from pick list		English
Provenance	description.provenance	Automatically added		
Publisher	publisher	Not used for theses		
Sponsors	description.sponsor	Use if research was sponsored by		
(sponsors and/or		an entity outside of the college		
funding codes)		which the thesis was submitted to		
Subject	subject	Use ProQuest list of keywords; list	690	Engineering
Keywords		each heading in a separate field		
Title	title	Use full formal title	245	Development of low temperature alumina coatings by AC magnetron sputtering
Type	type	Choose "Thesis" from pick list	502	Thesis
<i>Use the following fields only for theses</i>				
Advisor	contributor.advisor	List thesis advisor;	690	Matthew H. Gordon

(continued)

DSpace			Marc 21	
field name	Dublin Core name	Note	field	Example
		Use name authority record		
Committee	contributor.committee.	List all committee members	690	Joe Smith
Member	member	(if available) in separate fields		
<i>Use the following non Dublin Core fields only for theses</i>				
Degree Level	thesis.degree.level	Specify level of degree,	690	Undergraduate
		usually undergraduate		
Degree Name	thesis.degree.name	Specify brief name of the degree	690	Bachelor of Science
Department	thesis.degree.	department to which the paper	690	Mechanical Engineering
	department	was submitted		
Major	thesis.degree.discipline	Academic major of the student	690	Mechanical Engineering
Granting	thesis.degree.grantor	Institution which awarded the degree	690	University of Arkansas,
Institution				Fayetteville

Appendix B: MODS to MARC crosswalk for concert recordings

Field name	MODS field name	Note	Marc 21 field	Example(s)
Title	<titleInfo><title>	Object level metadata: the date of the recording in format YYYY-MM-DD Individual tracks: "[Track 01]", etc. Accompanying program: "Program"	Not mapped	<titleInfo> <title>2014-01-31</title> <subTitle> [Concert recording] </subTitle> </titleInfo> <titleInfo> <title> [Track 01] </title> <subTitle> [Concert recording] </subTitle> </titleInfo> <titleInfo> <title>Program</title> <subTitle> [Concert recording] </subTitle> </titleInfo>
Subtitle	<titleInfo><subTitle>	Always "[Concert recording]"	245 \$a	
Other Title	<titleInfo type="alternative"><title>	Alternate title of the performance, such as the actual title on the program, the name of the recital series, etc.	246 \$a	<titleInfo type="alternative"> <title>Wednesday Recital Hour</title> </titleInfo>

(continued)

Field name	MODS field name	Note	Marc 21 field	Example(s)
Name	<name type="corporate"><namePart>	Department name: "University of Arkansas, Fayetteville. Dept. of Music." Repeat the <namePart> element for each portion of the name	110 \$a \$b	<name type="corporate"><namePart>University of Arkansas, Fayetteville.</namePart><namePart>Dept. of Music</namePart></name>
Name	<name type="corporate"><namePart>	Name of ensemble(s). Repeat the <namePart> element for each portion of the name	710 \$a \$b	<name type="corporate"><namePart>University of Arkansas, Fayetteville.</namePart><namePart>Schola Cantorum</namePart></name>
Name Role	<name type="personal"><namePart><role><roleTerm>	Name of individual performer(s)	700 \$a \$e	<name type="personal"><namePart>Botes, Johan</namePart><role><roleTerm type="text">piano</roleTerm></role></name>
Name Role	<name type="personal"><namePart><role><roleTerm>	Name of conductor	700 \$a \$e	<name type="personal"><namePart>Payne, Jeff</namePart><role><roleTerm type="text">conductor</roleTerm></role></name>

Name Role	<name type="personal"> <namePart>Reed, Joshua</namePart> <role> <roleTerm type="text">composer<roleTerm> </role> </name>	700 \$a \$e	<name type="personal"> <namePart>Reed, Joshua</namePart> <role> <roleTerm type="text">composer<roleTerm> </role> </name>
Table of Contents	Program Notes. Format is: [Track 01]. Song Title / Composer -- [Track 02]. Song Title / Composer	505	<tableOfContents> [Track 01]. Rondeau / Jean Joseph Mouret -- [Track 02]. Dog gone blues / Luther Hendersen. <tableOfContents>
Date Issued	Date of performance (date recorded)	260 \$c, 518	<originInfo> <dateIssued keyDate="yes">2011-03-17 </dateIssued> </originInfo>
Subject	Type of recital	690	<subject> <topic>Saxophone recital</topic> </subject> <subject> <topic>Graduate recital</topic> </subject>
Type of Resource	Always "sound recording"	LDR 06	<typeOfResource>sound recording </ typeOfResource>

(continued)

Field name	MODS field name	Note	Marc 21 field	Example(s)
Note	<note>	Information about the recording, its contents, performers, etc., if deemed useful for access	500	<note>Track 12 contains both Yaravi and Froggy went a courtin' from Miniatures by William Grant Still.</note>
Location	<location><URL>	Islandora URL for the recording	856 \$u	<location> <url>http://hdl.handle.net/10975/islandora:193427</url> </location>
Location	<location><shelfLocator>	Local call number with CCD prefix, e.g., "CCD 140131"	099	<location> <shelfLocator>CCD 140131</shelfLocator> </location>