Notes on Operations
Metadata Revisited

Updating Metadata Profiles and Practices in a Vendor-Hosted Repository

A.L. Carson and Carol Ou

Implemented as a way to host open-access journals, the University of Nevada Las Vegas (UNLV) Libraries institutional repository (IR) expanded into collecting other researcher-created materials, a process that did not always include clear metadata and descriptive guidelines. Series-specific settings, unclear field definitions, and other varying practices created an inconsistent bibliographic database, however, and the unclear field definitions and lack of thorough internal documentation pointed to issues that would need to be addressed if the Libraries wanted to reliably share its IR metadata with its discovery layer and external harvesters and aggregators. To resolve this problem, UNLV undertook a metadata review intended to reconcile the fields used and provide recommendations on vocabularies and standards for capturing metadata. Through a collaborative, iterative process, the Metadata Review Team suggested and implemented changes to the IR's metadata structures, in consultation with vendor support, resulting in improved descriptive policies for IR resources.

The University of Nevada Las Vegas (UNLV) Libraries implemented its institutional repository (IR), Digital Scholarship@UNLV, on the Digital Commons platform in 2009. By 2016, it had become clear that the IR's ad-hoc approach to metadata standards had created an inconsistent bibliographic database, the result of series-specific settings, unclear field definitions, and varying practices over time. This irregularity became evident in a variety of ways, including when the Libraries' Discovery Services and Scholarly Communication Initiatives departments attempted to correct the mapping of metadata harvested from the IR in the Libraries' former discovery layer. The unclear field definitions and lack of thorough internal documentation pointed to issues that would need to be addressed if the Libraries wanted to reliably share its IR metadata both with its own discovery layer and specified external harvesters and aggregators.

The Libraries initiated a Libraries Fellows program in 2016, intended to provide new and early career librarians with "transferable professional early work experience and career development opportunities in preparation for future roles in the field."[1] The inaugural Fellows started at the beginning of 2017, and were assigned work across three project areas: research data management; scholarly research impact; and metadata support. The metadata support projects were intended to further the Libraries' goals in increasing the discoverability of UNLV's digital research outputs, which made a project to reconcile and document the Libraries' metadata practices in the IR a fitting assignment.

The assignment, dubbed a Metadata Review of the IR, sought to improve

A.L. Carson (acarson@iastate.edu) is Assistant Processing Archivist, Special Collections and University Archives, Iowa State University. **Carol Ou** (carol.ou@unlv.edu) is Head of Discovery Services, University of Nevada Las Vegas.

the reliability of metadata harvested from the IR by establishing clear field definitions, standardizing varying practices, reconciling those practices with the expectations of the Libraries' discovery layer and external harvesters and aggregators, and creating thorough documentation.

Literature Review

IRs present some specific complexities in relation to creating, managing, and sharing metadata. Chapman, Reynolds, and Shreeves's "Repository Metadata: Approaches and Challenges" describes the "mixed metadata environment" of IRs, featuring metadata from different sources and the resulting difficulty of "enforc[ing] consistent use of metadata and entry of metadata values."[2] This dynamic was clearly evident in Digital Scholarship@UNLV as it too had evolved over time with different series defining and using metadata fields differently in response to individual collection needs. Chapman, Reynolds, and Shreeves's paper also features case studies of three institutions that use the DSpace platform; the discussions repeatedly note DSpace's limitations and the workarounds each institution developed. Similarly, any changes to UNLV's IR would have to work within the structure of the Digital Commons platform.

The literature includes discussion of the utility of controlled subject vocabularies in IRs and their importance in supporting linked data. Hanrath and Radio's study of user search behavior supports the use of the FAST controlled subject vocabulary in the IR; they acknowledge the challenges of applying controlled subject vocabulary to IR content but also note that this work can be helpful in exposing repositories as linked data since FAST headings can be expressed as URIs.[3] Another paper by Radio and Hanrath discusses an actual effort to apply FAST headings to a subset of IR content followed by serializing the metadata into a linked data format. Using controlled subject terms was a fundamental part of their effort to expose a test set of records as linked data with the authors stating, "[linked data] records may also benefit from the consistency offered by use of a controlled vocabulary as necessitated by the use of unambiguous URI identifiers, particularly in contexts wherein such control had not previously been exercised."[4]

Sharing and repurposing metadata across contexts presents a significant opportunity and a not insignificant challenge for libraries. The Repository of Metadata Crosswalks offers solid context on the complications and mechanics of crosswalking, with a particular focus on crosswalking in an XML and web environment using applications of OAI.[5] Veve's "From Digital Commons to OCLC" specifically provides an example of harvesting and transforming metadata in the Digital Commons context, noting some of the particular challenges of Digital Commons' proprietary schema

and differences in metadata exposed via OAI-PMH.[6]

When considering this work from the aggregator's perspective, the "Guidelines for Encoding Bibliographic Citation Information in Dublin Core Metadata" and Potvin and Thompson's analysis of metadata standards to describe electronic theses and dissertations (ETDs) offer useful expectations for how the enhanced metadata should display to aggregators.[7] Sandy and Freeland's case study of ingesting and aggregating metadata from a group of institutions into the Digital Public Library of America (DPLA) notes issues such as "mismatches in data feeds from participating institutions" and the need to normalize the aggregated records using a Metadata Application Profile (MAP).[8] They further note the "importance of... local decisions supporting wide-scale interoperability."[9] Similarly, at UNLV, the Metadata Review of the IR was aware of the impact of local decision-making in creating and maintaining the metadata that would be shared and would seek internal consistency via the use of a MAP.

Background

UNLV is a public research university with an enrollment of twenty-nine thousand students, including approximately twenty-five thousand undergraduate and four thousand graduate students. The Libraries consist of one main library and four satellite libraries, employing more than 120 faculty and staff. The Libraries began implementing its IR on the Bepress Digital Commons platform in 2009, initially to host open-access journals and later as a more fully-fledged repository. Digital Commons is a fully hosted and vendor-supported system; Libraries staff can create and edit metadata and create new series in the system, but other types of changes require assistance from vendor support.

While the journals hosted through Digital Commons are managed by the university departments that publish them (content is uploaded to the platform by department staff, not Libraries staff), the bulk of materials housed in the IR are acquired, managed, and uploaded by Scholarly Communication Initiatives (SCI) staff. This includes the twice-yearly ingest of ETDs with metadata files (as acquired from ProQuest), plus faculty and other researcher output. SCI has workflows for inquiring, acquiring, and uploading faculty pre- and post-prints using Digital Commons' batch upload utility and metadata spreadsheet. User submissions are the exception, not the rule: since metadata capture is handled almost exclusively by Libraries staff, UNLV had a significant opportunity to establish uniform expectations for metadata fields in support of technical implementations. Vendor support could create or suppress metadata fields and adjust their mapping in the output, but input decisions such as how to format dates and which vocabulary to use

to populate a field could not be constrained at the software level; these gaps had to be addressed with policy, which could be developed and maintained by Libraries staff.

The Metadata Review listed a number of tasks and deliverables, including several laying the groundwork for future sharing of IR metadata with external harvesters and aggregators. These specific tasks included: reviewing current metadata practices, templates, and generated OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting) outputs from the IR; reconciling them with the metadata requirements for an initial list of desired external harvesters and aggregators; and comparing them against available best practices for metadata creation in IRs. The rationale was that doing this work prior to sharing IR metadata with any additional systems would ease metadata mapping decisions that would need to be made and reduce the need for future revisions. The deliverables initially included a documented MAP, and if necessary, a Metadata Reconciliation Plan. During the course of the assignment, it became clear that compiling a list of recommended changes, and working with Digital Commons support staff on implementing those changes, would also be a significant task and deliverable.

The review was intended to reconcile the metadata fields in use and provide recommendations on vocabularies and standards for capturing metadata, aligning with best practices in description and interoperability. These recommendations were drawn from work at similar institutions and in the area of repositories more generally, and when possible, reflected the IR's existing practices, focusing on making metadata capture and use consistent within collections and series. Recognizing that staffing levels, nature of materials, and numerous other factors influence the IR's daily operations, however, the scope of the review did not extend to proscriptive guidelines about levels of description per item or publication type, collection development or management policies, or other IR policies.

To provide context for this paper, two Libraries departments were involved in the Metadata Review: SCI and Discovery Services (DS). SCI manages and operates the IR, while DS is responsible for cataloging and related work that enables the discovery of library materials through the library catalog and discovery layer. The Fellow assigned to lead the metadata review was located in DS and worked closely with SCI staff on this project. When the project began, DS included a department head, three librarians, two classified staff, and the Library Fellow; for DS, the Library Fellow was the one primarily engaged with this project under the supervision of the department head. SCI included a department head, one librarian, one classified staff member, and another Library Fellow—all members of SCI served as stakeholders for this project.

The Metadata Review would also be affected by other systems changes occurring in the Libraries. The Metadata Review began in early 2017, and almost simultaneously, the Libraries were engaged in a migration to a new library system and discovery layer with a go-live date in December 2017. Specifically, the Libraries were migrating from the III (Innovative Interfaces, Inc.) Millennium integrated library system and ProQuest's Summon discovery layer to the Ex Libris Alma library services platform and Primo discovery layer. The change in discovery layer would directly touch on the Metadata Review as both the existing discovery layer (Summon) and the forthcoming discovery layer (Primo) needed to harvest metadata from the IR.

Environmental Scan

With a clear grasp of the goals and scope of the review, the next step was to survey the field. This survey had three aims: to understand current IR practices and capabilities; to identify aggregators' technical requirements; and to identify similar work that had already been done. Approaching the first two, in practice, collapsed into a gap analysis: the current state of the IR practices versus those necessary to support reliable sharing of IR metadata. The third goal, finding case studies about similar work, expanded to include material resources and other information to assist in bridging the gap defined in the course of the survey.

While SCI staff had a good understanding of their workflows, in reviewing the internal IR-management practices and support materials, the Fellow found that UNLV-specific documentation was out of date, cursory, or difficult to access. Vendor documentation for the Digital Commons product was more detailed and easily available, but the Fellow had questions about specific functionality (for instance, manipulating drop-down lists and exporting metadata items in specified formats) that required direct communication with support staff to resolve. One early finding of the review was that updating the documentation, both to reflect the changes made as part of the review and to make SCI practices transparent and consistent, was a major priority.

Following this review, the Fellow next addressed best practices for IR management, particularly concerning ETDs, such as the Networked Digital Library of Theses and Dissertations Interoperability Standard.[10] Drawing from the project brief, the Fellow compiled a list of harvesters and their metadata requirements beginning with Ex Libris' Primo and continuing with potential aggregators such as the Mountain West Digital Library (MWDL), the Association of Research Libraries' (ARL) SHARE, and OCLC's Digital Collections Gateway.[11] These aggregators headed the list because UNLV already shared digital collections information with MWDL (from a separate ContentDM instance) and Digital Collections Gateway,

although the latter was underused in part because of the lack of consistent metadata that made sharing IR materials difficult. A comparison of aggregator metadata schemes, requirements, and recommendations revealed many points of alignment (as in date formats or the use of the Dublin Core (DC) schema); where harvesters did not align, the Fellow sought out crosswalks or other supplementary materials to clarify what changes, outside of the Digital Commons environment, might be necessary to interact with those entities. The initial brief anticipated a relationship between UNLV and CrossRef for DOI creation, with IR metadata being transformed into the CrossRef schema and used to register DOIs. Instead, UNLV established a relationship with DataCite, necessitating a reevaluation of the metadata requirements based on DataCite's protocols.[12]

Reviewing the OAI-PMH protocol for harvesting metadata provided useful context for this project. OAI-PMH is one of the most common methods used to harvest local library metadata into discovery products.[13] The protocol itself is documented on the Open Archives Initiative site, described as "a low-barrier mechanism for repository interoperability," it defines both harvesters and repositories: harvesters are "operated by a service provider as a means of collecting metadata from repositories," while a "repository is managed by a data provider to expose metadata to harvesters."[14] For this project, UNLV's IR functioned as a repository and all the previously described harvesters and aggregators functioned as harvesters. The Open Archives Initiative site also lists implementation guidelines. Among the minimum requirements for repositories is the ability to output its metadata in the unqualified DC metadata format.[15] The Bepress Digital Commons platform's OAI-PMH implementation also supports qualified DC; this slightly richer metadata format had been previously selected by the Libraries as the preferred metadata format for harvesting and was the focus for any mapping improvements.

Method

Since the Metadata Review was managed by DS but would affect daily work in SCI, it was important to have a clear, methodical approach to define responsibility and keep the review on track. The goals of the review divided roughly into near-term (normalize and create clear guidelines for metadata capture in the form of MAPs), mid-range (improve discovery in our own systems and interoperability with outside systems), and long-term (position IR materials for sharing through linked data). Each goal required a set of changes; these changes ranged from structural adjustments in the IR, to documenting the new metadata capture procedures, to specific areas requiring remediation, an interconnected but distinct set of tasks shared between DS and SCI.

Work began with a list, created to help project stakeholders understand the steps needed to normalize metadata practices across the IR, of tasks that the Fellow would undertake to assess and adjust those practices. The first was a needs assessment, determining which publication types to prioritize while optimizing discovery and interoperability. The needs assessment took a fairly simple form: first, a content inventory of items by publication type; then, a survey of hit counts and download statistics to determine the most used collections. Using this information, the decision was made to privilege unique or distinctive works (such as the ETDs) over materials replicated elsewhere.

There are six publication types supported in Digital Commons: Series, Journal, ETD, Image Gallery, Community or Event, and Book. Each has its own needs, but while Digital Scholarship@UNLV includes all six types, it was clear that they would not all require the same level of improvement. The bulk of the material in the IR fell into the Series, ETD, or Journal publication type. As Journals in Digital Scholarship@UNLV are self-administered by the faculty or departments responsible for their creation, SCI staff were understandably hesitant to make changes that would affect the user experience for those administrators. Accordingly, the changes made to the Journal template chiefly addressed how metadata was outputted through OAI-PMH rather than how it was inputted through the self-deposit interface. Series and ETDs, however, are managed by SCI staff, granting the Fellow greater leeway when considering ways to improve the metadata capture practices in those high-use publication types.

Building on the material inventory and needs assessment, optimizing and adding fields was addressed next. Adding new fields and creating or strengthening usage guidelines supported the goal of making data consistent within and across publication types necessary for machine harvesting. For harvesting into the discovery layer, and potentially other systems, the IR relies on qualified DC records generated through internal mappings from Digital Commons metadata and exposed via OAI-PMH. Updated OAI-PMH mappings similarly required consistent and accurate use of the metadata fields; when reviewing the existing mappings revealed inconsistent or duplicative usage, those instances appeared in the list of recommended changes as points requiring clarification. The Fellow compiled suggestions for field and use changes into a recommendations list, which was then open for comment and discussion with DS and SCI stakeholders.

Suggested changes included adding a "Type" field to all publication types, mapping to "dc.type" in the OAI-PMH output, to clarify the nature or genre of the resource being cataloged. This change, which would enable better filtering in Bepress and those aggregators supporting it (including the library discovery layer), also served to capture

Figure 1. Portion of spreadsheet illustrating field and mapping changes and related harvester requirements.

preservation information for UNLV Libraries and satisfy metadata requirements for multiple potential aggregator partners. Proposed adjustments to how dates were captured and displayed focused on the ETD publication type, disambiguating between dates submitted for degree, degree awarded, and publication in the IR (which can vary due to embargoes), reflecting not only a need for greater clarity but also recommended best practices from the Networked Digital Library of Theses and Dissertations. Subject data capture was a high priority for the review, and the proposed changes to the subject fields are addressed in more depth below.

After an open comment period, the Fellow produced a Google Sheets workbook based on the batch update spreadsheets used by Digital Commons to update metadata and item records: this workbook contained seven sheets, one for each of the six publication types and one that provided information on how to read the spreadsheets. It presented a visual demonstration of how the suggested changes would look in the IR, what fields they would add, change, or remap, and how those fields could most logically be mapped or crosswalked to outside systems (see figure 1). With the changes laid out visually, it was easier to discuss in concrete terms what the changes would do and how they would affect the IR. Stakeholders, concerned about data loss, were anxious to establish that no existing fields would be removed as part of the changes: the recommendations called only for the addition of new ones and evaluation of current fields.

Combined with changed guidelines for values and usage, and updated OAI-PMH mapping, the recommendations were presented as lossless: metadata currently held in the IR would remain, awaiting remediation, but new records would be created according to the new guidelines.

The spreadsheet's layout helped everyone to see the proposed changes, how they would function, and how they would affect both local systems and external sharing. Changes to the metadata profiles took two forms: technical and procedural. The former dealt with changes in what fields would be included, how they would be named, how they would behave (in terms of allowed values and OAI-PMH mapping), and how they would appear in Digital Commons. The latter were changes that had to be made and implemented by SCI staff generating or capturing metadata as resources entered the repository: questions of usage, authorities, and often obligation (i.e., mandatory, recommended, optional) operate at the procedural level. While the first could be instantiated by contacting vendor support and requesting the changes, followed by testing the new profiles to validate their behavior, the second relied on discussion, documentation, and cooperation with SCI colleagues.

During the comment and discussion period, the issue of how to treat metadata-only records in the IR arose. The Digital Scholarship@UNLV Bibliography series collects citations for UNLV-affiliated work, showcasing and recording the output of UNLV scholars and researchers.[16] The resources in this series are typically record-only, linking to full-text versions outside the IR, which gave rise to an interesting question when discussing changes to the OAI-PMH mapping. Prior to the Metadata Review, these resources had been harvested to the discovery layer with the other IR content, appearing with other IR and library materials in the library catalog. The review prompted stakeholders to review this practice: the record-only resources did not represent items in the IR's collection, only links redirecting users elsewhere, so the question became whether it was appropriate to continue harvesting these resources to the Libraries' discovery system, and if not, what action to take. SCI and DS agreed that removing or suppressing the metadata-only items from the harvest made sense and that only items held or accessed through the Libraries should appear in the discovery layer. The Fellow consulted with Bepress support and proposed adding a field to the metadata structure, flagging whether OAI-PMH harvesting was enabled or disabled, accompanied by some guidance on how to use the field. This was a standard function available in Digital Commons that had not been previously used and proved to be a good solution to the problem of exposing metadata-only items to OAI-PMH harvesting.

Once the changes were presented in an actionable format, staff engaged in additional discussion about the effects these changes would have, both for DS and SCI. Some of the proposed changes, such as levels of obligation for given fields across publication types, were further revised following these discussions, resulting in a list of changes which, once enacted, would establish the new metadata structures for existing content and update the templates for future content. Rather than apply those changes to the entire IR at once, the departments agreed to test the new structures on a small sample of collections in each resource type, beginning with "Series," and working down the priority chain as determined by the appraisal. These test collections included both highly representative collections and edge cases (such as a collection of UNLV-produced podcasts about scholarly research on gambling), to test the fitness of the new profiles. The initial focus was on adding and testing the new fields: adjusting the OAI-PMH mapping was considered dependent on successful completion of the tests and the resolution of any issues arising from them.

The MAPs themselves took the form of a spreadsheet workbook, shared via Google Drive, with a page for each Publication Type plus a page explaining how to read the profiles. These were based on a small common set of mandatory elements (Title, Author, Date Published), with additional fields and obligation levels according to the needs of the materials and usefulness to users, both internal and external. In addition to specifying the fields in use for each Publication Type, the MAPs also specified (as much as possible) the format of the values to be entered. Since depositing into UNLV's IR is primarily done in batches by SCI staff rather than by researcher deposit through the user interface, it was possible to specify how metadata should be recorded even in those fields that Digital Commons could not feasibly restrict to a vocabulary or list. Instructions on how to record this information, which often bridged the gap between what Digital Commons could support on a software level and what aggregators required, were provided in the MAPs documentation, a Google Document accessible to everyone within the Libraries domain.

For an example of how the changes functioned and the specific problems the changes sought to address, it is useful to focus on the "dc.subject" field. In the IR's existing metadata profiles, subject information was contained in two fields: "keywords" and "disciplines." The latter refers to terms in the Bepress/Digital Commons' three tiered taxonomy of subject disciplines, while the former was used for both author-generated keywords and FAST headings (Faceted Application of Subject Terminology).[17] The practice of having controlled (FAST) and uncontrolled (author-generated) subject terms, undifferentiated, in a single subject field, complicated maintenance of FAST headings, making it difficult to distinguish controlled from uncontrolled terms in the DC output. This practice also presented a barrier for any future publication of IR metadata as linked data, as it

made it difficult to determine when and where URIs for FAST terms could be extracted. A "keyword" could be an intentionally assigned FAST heading, but could also be an uncontrolled keyword that happened to match a current FAST heading without actually containing the same meaning. To resolve the ambiguity, adding a controlled subject field to handle the FAST terms (already used in some collections) became a priority early in the review. Supporting this new field required changes to the OAI-PMH output, mapping it to dc.subject and removing the DC mapping for the uncontrolled subject field (testing the effect this would have on current records in the production setting is still ongoing).

In addition to the technical changes, DS recommended that FAST headings be applied whenever possible, recognizing that these would not always be relevant. The general recommendation for improving subject information was that one of the three subject fields (controlled, keyword, or disciplines) must contain a value: in this manner, "subject" is considered mandatory information in every record, but the type of subject information is not proscribed. This was believed to be the most balanced way to enrich description across all publication types without requiring information that might not exist. In the interest of making IR collections more discoverable in the broader Digital Commons Network (https://network.bepress.com/), DS recommended using the Bepress-controlled subject field (called "disciplines") whenever possible and at a minimum, for faculty publications and ETDs, where that information is readily available. The "disciplines" field is used within the Digital Commons Network to classify and make content discoverable across Digital Commons repositories; to raise the profile of Digital Scholarship@UNLV content in the Digital Commons Network it seemed to be in the interests of UNLV researchers and scholars, and the repository itself, to include this information.

The timing of these changes raised concerns about their effect on the production environment: at the time, the Libraries' discovery layer, Summon, harvested IR records weekly. Any alterations to the OAI-PMH mapping would necessitate corresponding adjustments to Summon, which would take two weeks to apply. Given that the Libraries were simultaneously navigating a migration to the Ex Libris Alma library services platform, with Primo serving as the future discovery layer, how much effort to expend on improving the mapping for an outgoing system was questioned. Additionally, Primo's harvesting and mapping configuration was entirely independent of the Summon configurations, meaning that any work done could not be repurposed for the new environment. Accordingly, the decision was made to focus on harvesting IR records and reconciling mapping changes in Primo.

In communication with Bepress support, collections for testing the field updates were identified in each of the

publication types; while small (averaging twenty items each), these collections included both highly-representative materials and those considered more unusual, to ensure the fitness of the new fields for the broadest possible application. Once the collections were specified, Bepress support added the new fields, notifying the Libraries when the changes were complete. At that point, SCI dedicated time to populating the new fields using the Digital Commons batch update spreadsheet, adding metadata to the test collections and allowing DS to check the output. Ensuring that metadata was structured as expected in the Bepress environment, mapped to DC as specified via the OAI-PMH output, and harvested correctly into Primo, a back-end process, comprised an end-to-end series of tests that needed to be completed prior to any additional work to adjust the public display of these records in Primo.

Testing the changes, once that data was added, was a two-pronged process: first, the Fellow used in-browser OAI-PMH calls to expose the qualified DC records created by Digital Commons, checking to ensure that the new fields appeared in the exported record, were mapped to the desired qualified DC fields, and contained the expected information (see figure 2). Second, searching for the records in the test series in Primo, the Fellow used the "display source record" function to see the qualified DC record as it had imported into Primo (see figure 3). In this manner, DS ensured that the changes behaved in the expected fashion, and that metadata in the new fields was expressed correctly in the Primo environment. When fields failed to appear or did not map to qualified DC as specified, the Fellow contacted Bepress support for a correction: only once this was complete could DS adjust how OAI-PMH data was displayed in the public-facing discovery interface. No testing can be considered complete without failure: one of the update requests did indeed result in a significant error, making OAI-PMH calls unresolvable for large sections of the IR collection. Fortunately, communication with Bepress support quickly resolved the issue, but it was an important illustration of what failure could occur.

Following the successful implementation and population of the new fields in test collections, during which the Fellow collaborated with SCI staff to develop and communicate metadata capture practices for the new fields, the next step was to formalize the changes to the MAPs. Instantiating the technical changes was straightforward: the Fellow contacted Bepress support to implement the new profiles to all collections in the six publication types. Template updates made by Bepress support ensured that all new collections created and new items added to existing collections would use the new MAPs: aligning existing collections with the new MAPs required some further communication, but was eventually completed. Ensuring consistent ongoing use of the new MAPs and procedural

```
 1  <?xml version="1.0" encoding="UTF-8"?>
 2  <OAI-PMH
 3      xmlns="http://www.openarchives.org/OAI/2.0/"
 4      xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
 5      xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/
 6          http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
 7      <responseDate>2018-10-11T20:26:39Z</responseDate>
 8      <request verb="GetRecord" identifier="oai:digitalscholarship.unlv.edu:brookings_pubs-1033" metadataPrefix="qdc" >https://digitalscholarship.unlv.edu/do/oai/</request>
 9
10      <GetRecord>
11
12  <record>
13      <header>
14          <identifier>oai:digitalscholarship.unlv.edu:brookings_pubs-1033</identifier>
15          <datestamp>2016-08-08T21:42:03Z</datestamp>
16              <setSpec>publication:mtnwest_pubs</setSpec>
17              <setSpec>publication:brookings_pubs</setSpec>
18              <setSpec>publication:brookings_capstone_studentpapers</setSpec>
19              <setSpec>publication:brookings_mtnwest</setSpec>
20      </header>
21      <metadata>
22          <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:schemaLocati
23
24          <dc:title>The End of the Road: The State of Urban Elevated Expressways in the United States</dc:title>
25          <dc:creator>Waqar, Daniel</dc:creator>
26          <dc:date.created>2016-05-01T07:00:00Z</dc:date.created>
27          <dc:publisher>Brookings Mountain West</dc:publisher>
28          <dc:description.abstract>&lt;p&gt;In a January 2016 meeting, the Clark County Commission heard a proposal from the County Public Works Department about a $200 million plan to buil
29          <dc:subject>Public Policy</dc:subject>
30          <dc:subject>Transportation</dc:subject>
31          <dc:subject>Urban Studies</dc:subject>
32          <dc:language>English</dc:language>
33          <dc:description>&lt;p&gt;This is a student research paper, patterned on a Brookings Policy Brief, written under the supervision of a Brookings Scholar.&lt;/p&gt;</dc:description>
34          <dc:identifier>https://digitalscholarship.unlv.edu/brookings_pubs/32</dc:identifier>
35          <dc:identifier>https://digitalscholarship.unlv.edu/cgi/viewcontent.cgi?article=1033&amp;amp;context=brookings_pubs</dc:identifier>
36          </oai_dc:dc>
37      </metadata>
38  </record>
39
40          </GetRecord>
41  </OAI-PMH>
42
```

Figure 2. Example OAI-PMH record generated by Bepress.

```
<?xml version="1.0" encoding="UTF-8"?>
<oai_dc:dc xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/" :
    <dc:title>The End of the Road: The State of Urban Elevated Expressways in the United States</dc:title>
    <dc:creator>Waqar, Daniel</dc:creator>
    <dc:type>Capstone Project</dc:type>
    <dc:date.created>2016-05-01T07:00:00Z</dc:date.created>
    <dc:publisher>Brookings Mountain West</dc:publisher>
    <dc:format.extent>1</dc:format.extent>
    <dc:format.extent>16</dc:format.extent>
    <dc:description.abstract>&lt;p&gt;In a January 2016 meeting, the Clark County Commission heard a proposal from
    <dc:subject>City planning</dc:subject>
    <dc:subject>Express highways</dc:subject>
    <dc:subject>Nevada--Las Vegas Metropolitan Area</dc:subject>
    <dc:subject>Urban transportation</dc:subject>
    <dc:subject>Public Policy</dc:subject>
    <dc:subject>Transportation</dc:subject>
    <dc:subject>Urban Studies</dc:subject>
    <dc:language>English</dc:language>
    <dc:description>&lt;p&gt;This is a student research paper, patterned on a Brookings Policy Brief, written unde
    <dc:identifier>http://digitalscholarship.unlv.edu/brookings_pubs/32</dc:identifier>
</oai_dc:dc>
```

Figure 3. OAI-PMH record imported into Primo.

changes (new expectations, metadata capture practices, and intended usages for the new fields, as well as revisions and clarifications of existing fields), however, would require user documentation.

Given the state of the pre-existing IR documentation and the broad nature of the changes, the Fellow determined that creating new documentation, with an outline of the project's intentions and decision logic to help guide future work, would be more informative and provide better context than attempting to update existing documentation. Accordingly, a document intended to cover the breadth of the review work was created: a narrative introduction to the project and its outcomes, a terms list defining IR-specific language used in the rest of the documentation, and a set of tables containing

definitions, instructions, and examples for the new MAPs. The bulk of the documentation is in these tables, which describe the fields and their purpose, and provide expected values for each, any relevant authorities for those values, and an example. Rather than reproduce information by annotating each MAP independently, the Fellow created a table of all those fields that are consistently used across the six publication types, with links out to smaller tables containing the fields or usages specific to that publication type. This document had some overlap with an existing citation formatting guide written (for an audience primarily of student workers) for the Digital Scholarship@UNLV Bibliography project: again, rather than reproduce work (and risk drift between the two documents), this common information was recorded

in another document and links to it were inserted in both the MAPs and the Bibliography documentation. By creating an interrelated corpus of documentation for IR practices, the Fellow hoped to not only connect all relevant information so that a user could access that information regardless of starting place, but make it easier to keep IR policy information current as practices changed.

Conclusion

The initial Metadata Review of the IR has been largely completed, and some next steps remain. Remediation and reconciliation of existing metadata to meet the minimal requirements of the new MAPs will be conducted as staff resources permit, likely prioritizing ETDs and other materials where the IR provides the full text. The OAI-PMH output with its revised and newly mapped elements will need to be tested against anticipated aggregators; to date, records from the IR have been harvested into Primo, the Libraries' new discovery layer, and testing will soon begin on using the OAI-PMH output as the metadata source for minting DOIs via DataCite. Lastly, the IR, and library metadata practices in general, do operate within a changing landscape; collection policies and system needs have evolved throughout the history of Digital Scholarship@ UNLV and will continue to do so.

It is therefore essential that metadata review activities do not function as occasional large projects but instead as part of the routine work of managing an IR. Moving forward, the departments will seek to build agility into this process so that metadata practices in the IR can be more responsive to changing expectations from aggregators and new developments in IR management, and documentation can keep pace with those developments, assisting the Libraries in maintaining institutional knowledge.

Reference

1. "UNLV Fellows Program," University of Nevada Las Vegas University Libraries, accessed October 16, 2018, https://www.library.unlv.edu/about/fellows.

2. John W. Chapman, David Reynolds, and Sarah A. Shreeves, "Repository Metadata: Approaches and Challenges," *Cataloging & Classification Quarterly* 47, no. 3–4 (2009), https://doi.org/10.1080/01639370902735020.

3. Scott Hanrath and Erik Radio, "User Search Terms and Controlled Subject Vocabularies in an Institutional Repository," *Library Hi Tech* 35, no. 3 (2017), https://doi.org/10.1108/LHT-11-2016-0133

4. Erik Radio and Scott Hanrath, "Measuring the Impact and Effectiveness of Transitioning to a Linked Data Vocabulary," *Journal of Library Metadata* 16, no. 2 (2016), https://doi.org/10.1080/19386389.2016.1215734.

5. Carol Jean Godby, Jeffrey A. Young, and Eric Childress, "A Repository of Metadata Crosswalks," *D-Lib Magazine* 10, no. 12 (2004), http://www.dlib.org/dlib/december04/godby/12godby.html.

6. Marielle Veve, "From Digital Commons to OCLC: A Tailored Approach for Harvesting and Transforming ETD Metadata into High-Quality Records," *Code4Lib Journal* no. 33 (2016), http://journal.code4lib.org/articles/11676.

7. Ann Apps, "Guidelines for Encoding Bibliographic Citation Information in Dublin Core Metadata," Dublin Core Metadata Initiative, accessed October 16, 2018, http://dublincore.org/documents/dc-citation-guidelines/; Sarah Potvin and Santi Thompson, "An Analysis of Evolving Metadata Influences, Standards, and Practices in Electronic Theses and Dissertations," *Library Resources & Technical Services* 60, no. 2 (2016), https://doi.org/10.5860/lrts.60n2.99.

8. Heather Moulaison Sandy and Chris Freeland, "The Importance of Interoperability: Lessons from the Digital Public Library of America," *International Information & Library Review* 48, no. 1 (2016), https://doi.org/10.1080/10572317.2016.1146041.

9. Ibid.

10. "Metadata," Networked Digital Library of Theses and Dissertations, accessed October 16, 2018, http://www.ndltd.org/standards/metadata.

11. "Overview of the Publishing Process," Ex Libris Knowledge Center, accessed October 16, 2018, https://knowledge.exlibrisgroup.com/Primo/Product_Documentation/System_Administration_Guide/010System_Architecture/040Overview_of_the_Publishing_Process; "The PNX Record," Ex Libris Knowledge Center, accessed October 16, 2018, https://knowledge.exlibrisgroup.com/Primo/Product_Documentation/Technical_Guide/010The_PNX_Record; "Best Practices for CONTENTdm and other OAI-PMH compliant repositories: creating sharable metadata Version 3.1," OCLC, accessed October 16, 2018, https://www.oclc.org/content/dam/support/wcdigitalcollectiongateway/MetadataBestPractices.pdf.

12. "DataCite Metadata Schema 4.1," DataCite Schema, accessed October 16, 2018, https://schema.datacite.org/meta/kernel-4.1/.

13. Edward M. Corrado, "Discovery Products and the Open Archives Initiative Protocol for Metadata Harvesting," *International Information & Library Review* 50, no. 1 (2018): 47–53, https://doi.org/10.1080/10572317.2017.1422905.

14. "Open Archives Initiative Protocol for Metadata Harvesting," Open Archives Initiative, accessed October 16, 2018, https://www.openarchives.org/pmh/; "Open Archives Initiative—Protocol for Metadata Harvesting—v.2.0," Open Archives Initiative, last modified January 8, 2015, https://www.openarchives.org/OAI/openarchivesprotocol.html.

15. "OAI-PMH Implementation Guidelines–Guidelines for Repository Implementers," Open Archives Initiative, last modified January 19, 2005, https://www.openarchives.org/OAI/2.0/guidelines-repository.htm.

16. "UNLV Author Bibliography," Digital Scholarship @UNLV, accessed October 16, 2018, https://digitalscholarship.unlv.edu/unlv_bibliography/.

17. "Digital Commons Three-Tiered List of Academic Disciplines (January 2017)," Bepress, https://www.bepress.com/wp-content/uploads/2016/12/Digital-Commons-Disciplines-taxonomy-2017-01.pdf.