Metadata Documentation Practices at ARL Institutional Repositories

Ayla Stein Kenfield

abstract: This paper describes an investigation into whether institutional repositories at member libraries of the Association of Research Libraries (ARL) provide publicly available metadata documentation via the repositories’ websites. The study compares the results of a September 2017 survey with findings from a review of ARL institutional repository websites. Slightly more than half of ARL institutional repositories do provide metadata documentation. Implications of the results are discussed, including specific research questions for moving forward. A call to reimagine documentation as a tool of transparency is delivered.

Introduction

Partially as a result of the advent of digital libraries and the growth of the open access movement, institutional repositories (IRs) have proliferated across academic libraries over the last 10 to 15 years. Metadata quality is an ubiquitous issue for institutional repositories, much as it is for other library systems. This paper focuses on the guidelines and documentation for metadata entry and formatting in the IRs of Association of Research Library (ARL) member libraries.

Institutional repositories indicate a broader change in library resource description. Unlike traditional library information retrieval systems, scholarly repositories commonly have materials described by people other than metadata or cataloging specialists due to creator self-archiving. In the self-deposit process, creators or their designated proxies submit a work directly into a repository, including providing descriptive information. Information
professionals will likely create metadata of higher quality than that developed by other classes of depositors due to their formal education and training, as well as their greater familiarity with local practices and policies. Research has shown, however, that providing metadata specialists and authors with guidelines for metadata creation results in higher quality metadata that are more useful to the local institution.¹

For the purposes of transparency and community oversight, providing metadata documentation is a metric for trustworthy digital and data repositories. While these standards and community best practices may fall into the realm of digital preservation rather than scholarly communication or resource description, they apply to IRs as well, despite much of the IR literature focusing on public facing-aspects. The trusted digital and data repository standards are applicable to discussions on IR metadata documentation because of the commitment to preserving the scholarly materials that libraries make by offering IR services. Additionally, the requirements enumerated in trusted digital and data repository standards force a certain transparency. Such clarity is crucial for cultural heritage and repositories in this era of black box algorithms and corporatized proprietary systems that are not open for inspection. Therefore, the questions posed in this paper are:

- Do institutional repositories at ARL libraries provide metadata documentation via the repository website?
- Will repositories with defined metadata policies make their metadata documentation available publicly?

To answer these questions, the author carried out a study comprised of two parts. First, a survey asked repository managers about the existence of metadata documentation for their IRs and how access is facilitated online. Second, a review of academic ARL institutional repository websites attempted to ascertain the existence of repository documentation.

**Literature Review**

There is little in the library literature dedicated to IR metadata quality issues. In 2003, Jane Barton, Sarah Currier, and Jessie Hey laid out several foundational research questions to explore metadata creation processes, as well as the importance and impact of quality assurance practices in metadata creation. Barton and her coauthors posited six key questions to guide future research. Two questions, in particular, are pertinent to this study: “What kinds of tools can be used to facilitate the metadata creation process, and how effective are they?” and “To what extent can the provision of guidelines, training, and support improve metadata creation?”² A year later, Marieke Guy, Andy Powell, and Michael Day presented a series of suggestions for achieving high-quality metadata in IRs, including the creation of “cataloging guidelines” which “should define all metadata elements that are to be used, give content guidelines for each—including information on the particular standards in use—and some examples.”³ Their article goes on to recommend that the guidelines be available as separate documentation, intertwined with data entry tools in some fashion, or a combination of both options. Guy and her coauthors did not indicate if they had tested the effectiveness of the strategies they recommended.⁴
Jung-ran Park found in reviewing literature about metadata quality that the provision of guidelines for metadata creation and data entry in digital repositories seemed to result in higher quality metadata. Park’s paper focused mostly on digital libraries, whose contents tend to be comprised of digitized special collection and archival materials, with minimal attention to the quality of IR metadata.

Providing further evidence of the benefits of metadata documentation, Mohammad Chuttur found that, overall, study participants who received more detailed metadata guidelines created records with fewer errors than those given only definitions for the elements, regardless of whether participants had experience or training in metadata creation. However, the guidelines had a greater effect on error reduction for some types of material than others.

A single article targeted administrative metadata specifically. Jane Otto surveyed ARL institutional repositories about metadata they considered administrative: digital preservation or provenance, technical data, and rights information. The results showed significant holes in data collection for all three areas and pointed to lack of staffing, complex standards, and diversity of workflows as potentially contributing factors. Despite requirements from digital preservation standards such as the Audit and Certification of Trustworthy Digital Repositories, Otto does not address the public availability of administrative metadata documentation or its collection.

Mary Kurtz analyzed the metadata records from three IRs with different approaches to educating contributors, providing metadata guidelines, and having librarians mediate and approve deposits. Kurtz found that it was not the repository with the most robust metadata documentation that had the most complete and accurate records, it was the one that required librarians to review submitted materials.

While trying to establish a best practice for determining IR policies and procedures, Steve Probets and Celia Jenkins examined the documentation from seven repositories and interviewed repository practitioners. Probets and Jenkins found that the documentation provided little, if any, information on metadata, let alone detailed policies or guidelines regarding metadata. Bijan Roy, Subal Biswas, and Parthasarathi Mukhopadhyay also reviewed policy documentation from a subset of members of the Coalition of Open Access Policy Institutions (COAPI), a group of universities and affiliated libraries that support making scholarly research freely available to the public. Roy, Biswas, and Mukhopadhyay found metadata policy and guidelines underrepresented in COAPI policy documents. In their paper, Roy and his coauthors do not differentiate between policy and metadata requirements or data entry guidelines.
Up until this point, the literature has mostly described research on the benefits of metadata guidelines for single-item deposits into a repository or for interoperability in high-level aggregations. In their 2016 paper, Ayla Stein, Kelly Applegate, and Seth Robbins found robust best practices and guidelines vital when ingesting collections and metadata in batch uploads. Even more crucial, however, is that the repository staff prioritize review and enforcement of the metadata policy and guidelines. There is little point in spending time and effort in the creation of rules if the repository administrators do not enforce them.13

In 2006, ARL published SPEC (Systems and Procedures Exchange Center) Kit 292: Institutional Repositories by Charles Bailey, Karen Coombs, Jill Emery, Anne Mitchell, Chris Morris, Spenser Simons, and Robert Wright, presenting the results of a survey on member libraries’ endeavors with IRs. The SPEC Kit is a high-level overview that highlights examples of documentation or design from a few repositories in each thematic section. One section presents files of metadata policies from the University of Kansas in Lawrence, the University of Texas at Austin, and the University of Utah in Salt Lake City, varying in breadth and detail of content. The universities of Kansas and Utah provided more comprehensive information than Texas’s example, including a data dictionary in Kansas’s case and a detailed handbook in Utah’s.14 However, none of the links provided are still functional.15

Alexandra Chassanoff administered a survey more narrowly focused on metadata quality assurance and procedures in ARL IRs. Although metadata documentation did not consist of the bulk of the survey, several questions were dedicated to the existence of staff and depositor guidelines, as well as whether respondents had policies specifying minimum metadata requirements. The results showed that a little over half of the institutions maintained some internal documentation, 72 percent provided depositor documentation, and 75 percent had policies specifying required metadata.16

Jung-ran Park and Yuji Tosaka surveyed digital library metadata creation operations hoping to learn more about the sharing of institutional metadata creation documentation. They argued that it was important for these records to be available publicly because they specified a baseline level for metadata quality.17 Taken together, these local metadata protocols could be used to create more consistent and interoperable metadata, or as guidance for institutions launching new services, as Lisa Gonzalez sought for representing serials metadata in a new IR.18 However, Park and Tosaka’s survey showed that sharing metadata creation documentation externally was the exception rather than the rule, supporting Karen Smith-Yoshimura’s results from several years prior.19

In 2015, Heather Lea Moulaison, Felicity Dykas, and Kristen Gallant distributed to randomly chosen repositories listed in the Directory of Open Access Repositories (OpenDOAR) a survey inquiring about metadata creation practices.20 The questionnaire centered on schema and controlled vocabulary usage, with some attention to metadata workers and best practices documentation. Primarily, Moulaison and her coauthors were interested in whether the metadata guidelines were developed locally or through the consensus of a community.21 It should be noted that OpenDOAR includes entries for any digital repository that hosts openly accessible materials and is not restricted only to IRs.

To the best of the author’s knowledge, a study on the provision of IR metadata documentation of the scope of this article has not yet been undertaken. This investigation
combines the review of nearly 100 ARL institutional repository websites for evidence of metadata documentation with a survey on metadata documentation practices of the same community. This research will provide necessary insight into a wide scope of metadata practices and transparency efforts at research libraries and into the relationship between supplying documentation and metadata quality.

Methods

Review of Institutional Repository Websites

Data collection began on August 30, 2016, and ran through June 7, 2017. Eligible institutions were identified from the 2017 list of ARL members, which were English-speaking or dual-language, degree-granting, academic research libraries that provided IR services. For 2017, there were 123 ARL members. Any library not affiliated with an university, such as independent research libraries and the Library of Congress, was removed. Five member institutions did not provide IRs and were also excluded. For the purposes of this project, all University of California campuses counted as a single institution. Ultimately, the list was reduced to 98 institutions.

Once the eligibility criteria were established, the website for each IR, if available, was inspected. The purpose of this scrutiny was to determine if metadata documentation for the IR was publicly available either on or linked from the repository website. The bar for what counted as metadata documentation was relatively low. If the metadata information at least specified recommended element names as well as their definitions, corresponding data entry rules, or other types of additional context, such as deposit form screenshots, then the author considered this information acceptable as “metadata documentation.” Metadata policies, submission instructions, metadata application profiles, local best practices documents, frequently asked questions, LibGuides, and similar documents could all be appropriate vehicles for displaying the metadata documentation. However, simply listing metadata labels without any definitions or additional context or only stating license permissions for how the repository metadata could be reused or harvested was not considered sufficient, even if the element names seemed self-explanatory to an information professional. Checking submission workflows for embedded metadata guidance was not feasible since most IRs require local university credentials to access; therefore, any documentation had to be publicly available.

The information sought in the website review was mostly determined prior to data collection. The author gathered descriptive data on the institution, library, and IR, such as the name and URL of each; existence and open availability of metadata policies and documentation (separately); type of metadata collected; and existence of recommended or required controlled vocabularies. Other data gathered were not thought of until the project had already begun, such as student enrollment and whether an institution was private or public. The website review data set is available from the University of Illinois, University Libraries, Illinois Data Bank, 2018, https://doi.org/10.13012/B2IDB-7323993_V1.

The metadata documentation had to be findable via the repository website in some way, either as native content or direct links. Indirect links—for example, a link from an
IR to a departmental website that presented information about the repository’s metadata on a child page—were also deemed acceptable. The key issue was whether users could locate the metadata documentation if they started on the repository website.

Survey

The second part of the project’s data collection consisted of a survey, Institutional Repository Metadata Documentation Practices at ARL Libraries, hereafter referred to as the survey. It asked respondents about the existence and provision of metadata documentation for their IRs. The survey was created and offered via Qualtrics software and featured 29 questions, including the informed consent agreement. During the design of the survey, the author met with a consultant from the University of Illinois Survey Research Lab for input on good survey design. She also solicited feedback from colleagues. Because the survey used skip logic, the next query a respondent saw might depend on the answer to a previous inquiry, so not every respondent answered every question.

Institutional repository librarians and those with similar or equivalent roles at ARL libraries were the target demographic; the survey requested that anyone outside this demographic did not participate. The complete eligibility language can be seen in Appendix A.

The survey was distributed via repository and scholarly communication e-mail lists not exclusive to ARL members. Additionally, the questionnaire went directly to IR librarians and departments that had been identified as lacking publicly available metadata documentation in the website review phase. The survey opened on Monday, August 28, 2017, and closed on Friday, September 29, 2017.

Results

Website Data Results

Demographics

In the course of this research, more data were collected than can be fully elaborated upon here. Therefore, this section will focus on a narrow set of questions related to the availability of metadata policies and metadata documentation, as separate ideas, on IR websites.

As mentioned, the author reviewed the IR websites of 98 ARL libraries that met the selection criteria; the list of institutions can be seen in Appendix B. As shown in Table 1, five eligible libraries did not provide an IR at the time of data collection. Additionally, two repositories were in development but not yet in production by the conclusion of data collection. Twenty-eight of the repositories were at private universities, and 70 were at public institutions.

Figure 1 shows the number of eligible institutions with IRs broken down by student enrollment ranges. Most of the eligible repositories are at universities with enrollments between 20,000 and 29,999 or between 30,000 and 39,999 students.
Table 1.
Numbers for website review

<table>
<thead>
<tr>
<th>Association of Research Libraries (ARL) members through calendar year 2017</th>
<th>123</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eligible ARL members</td>
<td>111</td>
</tr>
<tr>
<td>Reviewed ARL members with institutional repositories at time of data collection*</td>
<td>98</td>
</tr>
<tr>
<td>Eligible libraries without institutional repositories at time of data collection</td>
<td>5</td>
</tr>
<tr>
<td>Reviewed libraries with institutional repositories in development at time of data collection</td>
<td>2</td>
</tr>
<tr>
<td>Private ARL libraries with institutional repositories</td>
<td>28</td>
</tr>
<tr>
<td>Public ARL libraries with institutional repositories</td>
<td>70</td>
</tr>
</tbody>
</table>

*Data collected June 7, 2016, to August 14, 2017.

Figure 1. Institutions eligible for review of their institutional repository websites, by student enrollment.
Metadata Documentation and Policies

No community-accepted standard details what a metadata policy should entail or exclude. Some institutions have dedicated metadata policies for their digital collections, repositories, or catalogs that indicate how their metadata can be used or reused, along with the chosen license, preferred method of attribution, and similar requirements. Other organizations may specify the metadata elements available in their repository and which are required. At still other repositories, this information may not be included as part of a metadata policy but may appear in a different policy area, or it may not be mentioned at all. For these reasons, the author chose to look for the existence and availability of metadata policies separately from metadata guidelines or best practices documentation. They are distinct entities at the University of Illinois institutional repository IDEALS (Illinois Digital Environment for Access to Learning and Scholarship), with which the author is most familiar.

Regardless of the lack of consensus on the exact coverage of metadata policies, they remain important. Such policies often state how metadata are made available for harvesting, preferred attribution language, rights of the repository to enhance metadata, and so forth. OpenDOAR recommends that a repository’s metadata policy specify who can access the metadata and in what circumstances the metadata can be reused. A policy had to be a dedicated guideline for repository metadata to count as a metadata policy in this study. Analyzing the content of these policies was beyond the scope of this study but is an area of future research.

As Table 2 shows, the overwhelming majority of the reviewed repository websites, 72.45 percent, did not present distinct metadata policies. A handful, 11.22 percent of IRs, mentioned metadata in some fashion within other policies, such as guidelines on preservation, digital collections, library rights and roles, and withdrawal. Having a stand-alone metadata policy available on the IR website was the exception rather than the rule; only 16.33 percent of repositories did so. These results align with previous work examining the representation of metadata policies and documentation in IRs.

As mentioned previously, metadata documentation can take a variety of forms. To be considered in this study, the documentation had to provide a list of recommended metadata element names and some form of additional context.

Roughly 46 percent of reviewed IR websites did not provide publicly available metadata documentation. Fifty-four percent of them did, however, supply metadata documentation. The number of repositories exhibiting documentation for their metadata is significantly higher than the number that specify metadata policies.

Survey Results

At the close of the survey, there were 104 total responses, with 77 completed responses. A response was considered complete if a participant submitted answers and was directed to the “end of survey” confirmation page in Qualtrics. One additional response had to be thrown out during the analysis phase because the participant was not an ARL member, leaving 76 viable responses. This number dropped to 75 after one respondent declined to participate when answering the informed consent question. Finally, seven respondents indicated that their places of employment did not provide IR services, leaving a total of 68 answers for the remainder of the survey.
The majority of the survey respondents, 69 percent, were employed at public universities. Thirty-one percent worked at private institutions.

The author discarded a survey question that asked respondents whether their institution was minority-serving and, if so, which populations it served. Some respondents indicated their institution was minority-serving when in fact it was not.

Most institutions, 15, represented in the survey enrolled between 10,000 and 20,000 students, followed closely by 14 schools with enrollments below 10,000. Institutions enrolling between 20,000 and 30,000 students came in third highest with 12. Far fewer institutions enrolling between 40,000 and 50,000 students participated.

Following the demographics questions, the survey asked participants about levels of staffing dedicated to their IR. Table 3 shows the breakdown of answers from the question that asked respondents the number of individual full-time and part-time staff assigned specifically to the IR. The vast majority of respondents indicated that their repository had 0 to 2 dedicated full-time equivalent staff, which is consistent with anecdotal evidence. Three respondents reported that they had five or more full-time equivalents assigned to their repository. Most of the respondents’ institutions dedicate both full-time and part-time staff to the IR.

After staffing inquiries, a series of questions asked survey participants about repository administration. The first question in this set inquired, “How is your repository designed to be managed? (May be different than how it is operated de facto).” The results are shown in Table 4.

The leading management methods under which IRs operated were distributed management, centralized management, and both. Distributed management means that each community or equivalent has a designated administrator who handles the daily management of the community, which may include (but is not limited to) approving individuals to submit materials to their community; reviewing and approving submitted items, metadata, or both; creating collections within the community; performing outreach to the community’s stakeholders; and depositing materials on behalf of researchers. In centralized management, a single top-level administrator or administrative team handles management tasks for all communities across the repository. These tasks may include (but are not limited to) approving individuals to submit materials for all communities.

### Table 2.
Metadata policies in institutional repositories

<table>
<thead>
<tr>
<th>Metadata policy status</th>
<th>Number</th>
<th>Percentage of institutional repositories</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>71</td>
<td>72.45%</td>
</tr>
<tr>
<td>Mentioned in other policies</td>
<td>11</td>
<td>11.22%</td>
</tr>
<tr>
<td>Dedicated policy</td>
<td>16</td>
<td>16.33%</td>
</tr>
</tbody>
</table>
Table 3.
Survey results regarding dedicated institutional repository staff

<table>
<thead>
<tr>
<th>Full-time</th>
<th>Part-time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Management design type</td>
<td>Percentage</td>
</tr>
<tr>
<td>Distributed*</td>
<td>76.92%</td>
</tr>
<tr>
<td>Centralized†</td>
<td>18.46%</td>
</tr>
<tr>
<td>Both ‡</td>
<td>4.62%</td>
</tr>
<tr>
<td>Total</td>
<td>100%</td>
</tr>
</tbody>
</table>

* Distributed management means each community has a designated administrator who handles the daily management, which may include approving individuals to submit materials to their community; reviewing and approving submitted items, metadata, or both; creating collections within the community; performing outreach to the community’s stakeholders; and depositing materials on behalf of researchers.

† Centralized management means a single top-level administrator or administrative team handles administrative tasks for all communities. These tasks may include approving individuals to submit materials for all communities or collections; reviewing and approving deposits into the repository (both individual submissions and batch uploads); creating new communities; creating new collections within existing communities; and performing outreach to repository stakeholders.

‡ Both means the repository is designed to have designated administrators for specific communities as well as centralized management across the repository’s communities.

Table 4.
Management design of institutional repositories

<table>
<thead>
<tr>
<th>Management design type</th>
<th>Percentage</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distributed*</td>
<td>4.62%</td>
<td>3</td>
</tr>
<tr>
<td>Centralized†</td>
<td>36.92%</td>
<td>24</td>
</tr>
<tr>
<td>Both ‡</td>
<td>58.42%</td>
<td>38</td>
</tr>
<tr>
<td>Total</td>
<td>100%</td>
<td>65</td>
</tr>
</tbody>
</table>

* Distributed management means each community has a designated administrator who handles the daily management, which may include approving individuals to submit materials to their community; reviewing and approving submitted items, metadata, or both; creating collections within the community; performing outreach to the community’s stakeholders; and depositing materials on behalf of researchers.

† Centralized management means a single top-level administrator or administrative team handles administrative tasks for all communities. These tasks may include approving individuals to submit materials for all communities or collections; reviewing and approving deposits into the repository (both individual submissions and batch uploads); creating new communities; creating new collections within existing communities; and performing outreach to repository stakeholders.

‡ Both means the repository is designed to have designated administrators for specific communities as well as centralized management across the repository’s communities.
or collections; reviewing and approving deposits into the repository (both individual submissions and batch uploads); creating new communities; creating new collections within existing communities; and performing outreach to repository stakeholders. A repository employing both management methods has designated administrators for specific communities as well as centralized management across all the repository communities.

Over half of respondents indicated that their institutional repository was managed both centrally by a repository administrator and in a distributed manner with community or collection administrators. Repositories run centrally were the second highest, with a little over a third of participants choosing this answer. Systems designed for distributed management made up only 4.62 percent of repositories.

Table 5 shows the results of the question “How are materials deposited into your institutional repository?” The survey asked participants to select from three methods, defined as the following:

1. Unmediated: Any permitted user can submit individual scholarly works to the institutional repository. There is no repository-level requirement that all materials and metadata be reviewed before final acceptance into the repository.
2. Mediated self-deposit: Any permitted user can submit individual scholarly works to the institutional repository. There is a repository-level requirement that all materials and metadata be reviewed by repository staff before final acceptance into the repository.
3. Repository staff-proxy deposit: No users are permitted to submit individual scholarly works to the institutional repository; repository staff handle submission of all materials. Researchers must send their files to repository staff for their work to be added to the repository.

Originally, 21.54 percent of participants responded “Other.” However, all the free-text responses were a mix of the named options, for example, both “mediated and unmediated” or “mediated, unmediated, and staff.” Therefore, five additional categories were created, indicating specific mixtures of deposit options, for example, “Staff uploads + unmediated,” and the free-text responses were redistributed accordingly. Additionally, the option to deposit via a current researcher information system (CRIS)—that is, a system that maintains information about research conducted at an institution—was left as an option on its own. The most frequent response was “Mediated self-deposit,” where every submission and its metadata are reviewed prior to ingest.

Repository staff-proxy deposit was the second most frequent response because this option typically results in higher-quality metadata and reduces pressure from researchers to self-archive their work. Unmediated self-deposit outnumbered the “Staff-upload + mediated” combination category, with nine respondents selecting this answer. This method, in theory, balances ease-of-deposit and repository staff time while potentially sacrificing metadata quality.

The next group of survey questions focused on the IR metadata documentation, the main point of the survey. The first question in this set asked, “Does your IR provide documentation and/or guidance on your IR’s metadata?” The 65 responses to this question were nearly evenly split. Almost 54 percent of participants said their repository provided metadata documentation, while 46 percent indicated that there was no metadata documentation for their organization’s IR. The author expected that result because leaving the determination of what to consider metadata documentation to the survey participants meant that far more IRs would be reported to supply this information.
Librarians, especially those within scholarly communication and research data management domains, routinely emphasize—for the purposes of peer analysis, transparency, and research reproducibility—the need for researchers to document their data, create README files that provide information for the user about any software programs developed, and deposit their work tidily into a repository at the conclusion of a research project. However, librarians do not always follow our own best practices. Otherwise, the percentage of IRs providing metadata documentation would be much higher than 54 percent.

Table 5.
Answers to the survey question “How are materials deposited into your institutional repository?”

<table>
<thead>
<tr>
<th>Deposit options</th>
<th>Number</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unmediated self-deposit*</td>
<td>9</td>
<td>13.64%</td>
</tr>
<tr>
<td>Mediated self-deposit†</td>
<td>26</td>
<td>39.39%</td>
</tr>
<tr>
<td>Repository staff-proxy deposit‡</td>
<td>17</td>
<td>25.76%</td>
</tr>
<tr>
<td>Mediated and unmediated</td>
<td>3</td>
<td>4.55%</td>
</tr>
<tr>
<td>Mediated, unmediated, and staff uploads</td>
<td>3</td>
<td>4.55%</td>
</tr>
<tr>
<td>Staff uploads and unmediated</td>
<td>2</td>
<td>3.03%</td>
</tr>
<tr>
<td>Staff uploads and mediated</td>
<td>5</td>
<td>7.58%</td>
</tr>
<tr>
<td>CRIS deposit§</td>
<td>1</td>
<td>1.52%</td>
</tr>
<tr>
<td>Total</td>
<td>66</td>
<td>100%*</td>
</tr>
</tbody>
</table>

* Unmediated self-deposit means any permitted user can submit scholarly works to the institutional repository. There is no repository-level requirement that all materials and metadata be reviewed before final acceptance into the repository.
† Mediated self-deposit means any permitted user can submit scholarly works to the institutional repository. There is a repository-level requirement that repository staff review all materials and metadata before final acceptance into the repository.
‡ Repository staff-proxy deposit means no users are permitted to submit scholarly works to the institutional repository; repository staff handle submission of all materials. Researchers must send their files to staff for them to be added to the repository.
§ CRIS (current researcher information system) is a database or other information system used to store and manage data about research conducted at an institution.
* Rounded.
the percentage of IRs providing metadata documentation would be much higher than 54 percent.

The question “Does the metadata documentation/guidance provided indicate the required metadata elements?” was only presented to those who answered “Yes” to providing metadata documentation. Reassuringly, 82.86 percent of respondents to this question said that the documentation did indicate required metadata.

Participants who answered that their metadata documentation did not indicate required metadata elements were directed to a question asking why this information was not covered. Eleven respondents answered, as shown in Table 6. The two major reasons why this information was not included in the documentation were that required elements were indicated in the submission workflow and that repository staff handled metadata creation. Additionally, several respondents commented that their user community had not requested this information. It was surprising that staff creation of metadata was so common a reason for not indicating required metadata elements because even staff need this information. Further, it can help patrons ensure they provide what staff need to accurately describe the materials.

The final pair of questions asked survey participants if their IR documentation, not necessarily metadata-exclusive, included data entry rules. Most respondents, 57.14 percent, integrate data entry rules with non-metadata documentation, such as submission instructions. Providing stand-alone data entry information was the next highest response at 20 percent, followed by no data input documentation at all, at 17.14 percent. Only two respondents indicated that all metadata-related information was provided together. See Table 7.

Those who answered that metadata input guidelines were not supplied were directed to a follow-up “select all that apply” inquiry asking why this information was not included. There were only six responses, four of which were free-text “Other” answers. One participant expressed doubt that any data entry rules would be followed even if documentation was proffered. Another wrote that the repository staff assisted depositors with personal guidance but did not provide documentation. This response, in particular, highlighted a terminology flaw in the survey: while the author used documentation and guidance interchangeably, this respondent apparently considered guidance to be advice provided verbally rather than in writing. Still another free-text response said data entry documentation was not supplied because the requirements varied depending on the type of material. The remaining reply stated that data entry documentation was not provided for the public because any user-facing system requiring documentation reflected poor user-interface design, but that the IR did maintain staff-only documentation. See Table 8.

Table 9 shows responses to the question “How is the institutional repository metadata documentation made available to users?” A plurality of respondents, nearly
Table 6.
Why required metadata elements are not covered by documentation in institutional repositories

<table>
<thead>
<tr>
<th>Requirement</th>
<th>Percentage</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Required metadata elements are indicated in submission workflow.</td>
<td>36.36%</td>
<td>4</td>
</tr>
<tr>
<td>Required metadata are pulled from file properties.</td>
<td>9.09%</td>
<td>1</td>
</tr>
<tr>
<td>Not enough staff to create documentation.</td>
<td>0.00%</td>
<td>0</td>
</tr>
<tr>
<td>Documentation is in development.</td>
<td>9.09%</td>
<td>1</td>
</tr>
<tr>
<td>Metadata creation is handled by staff.</td>
<td>36.36%</td>
<td>4</td>
</tr>
<tr>
<td>Other reasons, such as more public documentation public for electronic theses and dissertations than anything else; most metadata standards for internal use only.</td>
<td>9.09%</td>
<td>1</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>100%</strong></td>
<td><strong>11</strong></td>
</tr>
</tbody>
</table>

Table 7.
Data entry rules in institutional repositories

<table>
<thead>
<tr>
<th>Rule</th>
<th>Number</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metadata input documentation/guidance is integrated with other (non-metadata) documentation (for example, the “how to submit an item to the repository” documentation).</td>
<td>20</td>
<td>57.14%</td>
</tr>
<tr>
<td>Documentation/guidance for metadata entry is provided as stand-alone data entry rules and is not incorporated into other documentation.</td>
<td>7</td>
<td>20.00%</td>
</tr>
<tr>
<td>All institutional repository metadata information is provided together.</td>
<td>2</td>
<td>5.71%</td>
</tr>
<tr>
<td>Metadata input documentation/guidance is not provided.</td>
<td>6</td>
<td>17.14%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>35</strong></td>
<td><strong>100%</strong></td>
</tr>
</tbody>
</table>
Table 8.
Why institutional repositories do not provide documentation of data entry rules

<table>
<thead>
<tr>
<th>Reason</th>
<th>Percentage</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data entry rules documentation is on the to-do list but is not a high priority.</td>
<td>0.00%</td>
<td>0</td>
</tr>
<tr>
<td>There are no free-text metadata fields.</td>
<td>16.67%</td>
<td>1</td>
</tr>
<tr>
<td>Not enough staff.</td>
<td>0.00%</td>
<td>0</td>
</tr>
<tr>
<td>Documentation is in development.</td>
<td>16.67%</td>
<td>1</td>
</tr>
<tr>
<td>Other</td>
<td>66.67%</td>
<td>4</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>100%</strong></td>
<td><strong>6</strong></td>
</tr>
</tbody>
</table>

half, make their metadata documentation available publicly. Staff-use only metadata documentation was the second highest response, with “Available by request” following. Three free-text responses were a mix of “Publicly available” and “Log-in required,” the exact meanings of which are ambiguous. The documentation may be integrated into the submission workflow, or there may be both public and staff-specific versions of the documentation. Two respondents indicated that their documentation is only available with a log-in. Based on the free-text responses, this question might have worked better as a “select all that apply” inquiry. The entire survey text is available in Appendix A.

Comparison of Survey and Website Results

Table 10 compares numbers from the website review with the survey data. There were 20 more repository websites reviewed than there were survey respondents. However, the percentages of ARL institutional repositories and survey participants with IR services were close, 88.29 percent and 90.67 percent, respectively. The same was also true for the percentages of private and public institutions, which differed by less than 2.5 percent between the survey and website data.

The question at heart of this research is whether documentation on the metadata requirements or best practices for an IR is offered and if the documentation is available on the repository website, either as a link to an external object or as native content. Roughly 46 percent of ARL institutional repository websites lack metadata documentation that the author could identify and locate.
Table 9.
Availability of metadata documentation in institutional repositories

<table>
<thead>
<tr>
<th></th>
<th>Percentage</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Publicly available</td>
<td>48.57%</td>
<td>17</td>
</tr>
<tr>
<td>Available by request</td>
<td>11.43%</td>
<td>4</td>
</tr>
<tr>
<td>Log-in required</td>
<td>5.71%</td>
<td>2</td>
</tr>
<tr>
<td>Metadata documentation for staff use only</td>
<td>17.14%</td>
<td>6</td>
</tr>
<tr>
<td>Combination of publicly available and log-in required</td>
<td>8.33%</td>
<td>3</td>
</tr>
<tr>
<td>Combination of publicly available and documentation for staff use only</td>
<td>2.78%</td>
<td>1</td>
</tr>
<tr>
<td>In development</td>
<td>2.78%</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>100%</td>
<td>35</td>
</tr>
</tbody>
</table>

and locate. However, the majority did provide some kind of metadata documentation, as stated previously. The survey result numbers align closely with the website review numbers. Nearly 54 percent of survey respondents indicated that metadata guideline documentation is provided for their IR, while 46 percent of respondents said metadata documentation was not supplied, which is nearly identical to the website review results.

Of the 53 IR websites that provide documentation, 38, or about 71.7 percent, recommend or require specific data entry rules, controlled vocabularies, or both. This number is encouraging because research has shown that when metadata creators have documentation that includes data entry guidelines, their records have fewer errors than those created without guidelines or with documentation that states only the metadata field definitions.

Generally, more of the metadata documentation found by the researcher contained information about data entry rules than provided controlled vocabularies. Indeed, terms selected from drop-down menus may have been properties in a standardized controlled vocabulary, and if not formally presented as such, curated options in a drop-down list still constitute a local controlled vocabulary, regardless if they originated in a formal standard. However, only a handful of IRs specified their recommended or required controlled vocabularies in their metadata documentation. Additionally, when a drop-down menu was mentioned, the options in the menu were not specified in
the documentation. Of course, some content entry rules and controlled vocabularies are incorporated in repository software, especially out-of-the-box systems such as DSpace.\textsuperscript{32} Far more of these repositories may use controlled vocabularies and enforce data entry rules than the 38 found here, but by-and-large, this information was not discernible from available metadata documentation.

Table 11 shows how the respondents answered the survey’s corresponding question on data entry rules. Thirty-five participants answered. Much like the results depicted in the website data, the majority of these respondents, 82.85 percent, selected one of the three possible “yes” answers, stating that data entry information was indeed included in their IR metadata documentation. The survey answer

<table>
<thead>
<tr>
<th>Website review question</th>
<th>Website review question value</th>
<th>Website review percentage</th>
<th>Survey question value</th>
<th>Survey response percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of eligible members of Association of Research Libraries (ARL)</td>
<td>111 participants</td>
<td>N/A</td>
<td>Agreed to</td>
<td>75</td>
</tr>
<tr>
<td>Number of reviewed ARL members with institutional repositories at time of data collection</td>
<td>98</td>
<td>88.29%</td>
<td>Answered “Does your institution provide an institutional repository?” yes</td>
<td>90.67%</td>
</tr>
<tr>
<td>Number of eligible libraries without institutional repositories at time of data collection</td>
<td>5</td>
<td>4.5%</td>
<td>Answered “Does your institution provide an institutional repository?” no</td>
<td>9.33%</td>
</tr>
<tr>
<td>Number of private ARL libraries with institutional repositories</td>
<td>28</td>
<td>28.57%</td>
<td>Answered “Is your institution public or private?” private</td>
<td>30.77%</td>
</tr>
<tr>
<td>Number of public ARL libraries with institutional repositories</td>
<td>71.42%</td>
<td>Answered “Is your institution public or private?” public</td>
<td>69.23%</td>
<td></td>
</tr>
</tbody>
</table>

Only a handful of IRs specified their recommended or required controlled vocabularies in their metadata documentation.
The overwhelming majority of respondents said that the information on data entry is integrated into non-metadata specific documentation. In fact, submission instructions were also the most popular form of data entry rules found on the reviewed repository websites. A mere handful of survey responses stated that their data entry rules existed as dedicated stand-alone documentation. More survey participants said that data entry information was not provided for their repository at all than those who consolidated all their metadata guidelines into a single document.

This research found that over half of ARL members’ IR websites present, link to, or otherwise make available metadata documentation. Fifty-four percent of the reviewed websites did so, and 53.85 percent of the survey respondents concurred. The proximity of the resulting percentages suggests that the data scoured from the repository websites reasonably reflect reality; a low level of response bias in the survey results can also be inferred. Assuming the participants read and followed the directions specifying survey eligibility, all the respondents’ institutions would be contained within the larger circle of the reviewed repositories. Of course, an examination of all IR websites at English-speaking North American higher education institutions might show similar rates of metadata documentation provision, but a project of that nature is out of scope for this paper.

This research also provides evidence supporting a previous study that found “generally repositories of elite institutions have more detailed policy documentations [sic] than those [open access repositories] that are at implementation stages or in their infancy.” ARL libraries, which tend to be better funded than their counterparts, may more likely have the resources to create metadata guidelines for their IRs.

Table 11. Data entry in institutional repositories

<table>
<thead>
<tr>
<th>Number</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metadata input documentation/guidance is integrated with other (non-metadata) documentation (for example, the “how to submit an item to the repository” documentation).</td>
<td>20</td>
</tr>
<tr>
<td>Documentation/guidance for metadata entry is provided as stand-alone data entry rules and is not incorporated into other documentation.</td>
<td>7</td>
</tr>
<tr>
<td>All institutional repository metadata information is provided together.</td>
<td>2</td>
</tr>
<tr>
<td>Metadata input documentation/guidance is not supplied.</td>
<td>6</td>
</tr>
<tr>
<td>Total</td>
<td>35</td>
</tr>
</tbody>
</table>
The author hypothesized that, if an IR had a metadata policy, it would likely make metadata documentation available on its website. This was not the case. Institutions that lacked metadata policies were statistically just as likely to provide metadata documentation as not. The content of the metadata policies and documentation was not analyzed for differences; information the author would consider metadata documentation may have been incorporated into metadata policies.

In the majority of cases, documentation on repository metadata was available on IR websites, but just under half of repositories reviewed and surveyed did not provide documentation. The lack of documentation may result from shortages of staff and time, project prioritization, or a plethora of other reasons. One survey respondent, taking a user experience design standpoint, stated, “Depositing authors should not require instructions, or (our feeling is), the system is not usable!”

Quality assurance tools that validate metadata field values, drop-down lists, and semiautomatic or auto-generated metadata are essential. Such aids should be integrated into IR metadata creation tools and submission workflows to ensure as painless, and accurate, a deposit process as possible. Indeed, Park found that incorporating a “simplified version of the metadata guidelines” into Web forms improved the quality of metadata.

In contrast to the results presented here, Park and Tosaka reported that “the vast majority of institutions (72.0 percent) provided no public access to local application profiles on their websites while only 19.6 percent of respondents’ institutions made them available online to the public.” Their focus was on metadata guidelines for digitized special or archival collections rather than for IRs. Park and Tosaka’s findings could result from differences in practitioner philosophies, since IRs were devised as a way to facilitate open access and combat entrenched academic publishing systems. The disparity in results could also be attributed to Park and Tosaka not limiting their study to ARL institutions. Attitudes toward sharing metadata guidelines publicly may also have changed in the nine years since their investigation.

Harking back to Chassanoff’s survey, she found that a slim majority of respondents, limited to ARL institutional repositories, maintained internal requirements for metadata. Seventy-two percent in Chassanoff’s survey provided metadata guidelines for depositors, a larger proportion than found in this study or in Park and Tosaka’s work. The high percentage of institutions that provided metadata guidelines to depositors in Chassanoff’s study appears at first to support the notion that IRs are founded on fundamentally different values toward access than those underlying digitized special collections materials; however, IRs rely more heavily on depositor-created metadata than traditional digital libraries do.

Strangely, the existence of metadata policies for ARL institutional repositories has decreased significantly since Chassanoff’s study. The website review found metadata
In a time of ever more interconnectivity, who should be considered a “user” of IRs must be broadened.

71.21 percent of all ARL institutional repository programs rely on submitter-created metadata in some fashion.

metadata policies for, at best, 27.55 percent of reviewed IRs (11.22 percent mentioned metadata in related policies, and 16.33 percent had stand-alone metadata policies), a far lower percentage than in Chassanoff’s study, where 75 percent of respondents reported that their institution had metadata policies.38

The intended purpose of metadata guidelines and documentation needs to be expanded. For example, in addition to meeting the typical needs of a depositor, does the available metadata documentation try to anticipate the needs of future colleagues? Does it support harvesting by metadata aggregation services? Does it meet the needs of scholars whose research entails text-mining a repository’s metadata? Is the presentation of metadata documentation approachable not only for experienced users and librarians but also for newcomers to the library and information science profession? Metadata schema, MARC (MAchine-Readable Cataloging) standards, and controlled vocabulary documentation are often nearly indecipherable to novices in the field as well as to experienced information professionals. In a time of ever more interconnectivity, who should be considered a “user” of IRs must be broadened. For instance, repository administrators are a type of user as well as self-archiving researchers.

Research data management and scholarly communication librarians need to lead by example. In workshops and consultations, researchers are encouraged to identify a data repository before beginning their research so the information requested by the repository can be built into the data collection procedures from the start. It is contradictory to advocate that researchers extensively document their research data, software code, information systems, methods, and other information when library scholarly repository services do not reflect the best practices information professionals teach.

Institutional repositories are traditionally document-centered collections, often with familiar bibliographic metadata. In response to the question “How are materials deposited into your institutional repository?” a plurality of respondents, 39.39 percent, said their collection methods were primarily “mediated self-deposit,” where researchers submit their work and library staff review it prior to final approval and ingest. If all categories that include any kind of depositor-provided metadata are counted together, this plurality becomes the majority, and 71.21 percent of all ARL institutional repository programs rely on submitter-created metadata in some fashion. Although the descriptive information typically asked for in deposit workflows seems obvious to library and archives professionals, it is exclusionary to assume knowledge or to presume that the information asked for is clear to all users.

One influential framework for representing IR metadata has been the 2006 article “Moving towards Shareable Metadata” by Sarah Shreeves, Jenn Riley, and Liz Milewicz.39 The FAIR Data Principles, also known as the Fair Guiding Principles, formally published in 2016 as the FAIR Guiding Principles, present a new paradigm to make research data “Findable, Accessible, Interoperable, and Reusable.”40 Why should IR
librarians care about a framework for data repositories? The answer is threefold. First, the FAIR Principles specify that they apply to both research data and metadata. The second reason is the blurring lines that separate data repositories, digital libraries, and IRs. Increasingly, research libraries choose to provide a unified digital repository that hosts all genres of digitized and born-digital content and data rather than maintaining multiple, disparate systems. As a result, the realms of research data publication and provision of open access scholarship are ever more intertwined. The FAIR Principles build on the legacy set by Shreeves and her colleagues to provide an updated framework that has the ability to benefit multiple kinds of research and scholarly outputs.

The final reason why IR librarians should care about the FAIR Principles is transparency and trust. For centuries, libraries and archives across societies and cultures have been entrusted with collecting and preserving creative works and scholarship. In the current cultural moments of “post-truth” and “alternative facts,” transparency of practice is essential for maintaining societal trust. Preservation and stewardship standards such as TDR (trustworthy digital repositories) or the requirements for CoreTrustSeal, which certifies data repositories as trusted digital repositories, require provision of metadata documentation for transparency so that a repository’s designated community can verify the authenticity of its collections.

While high quality and interoperability are desirable, if not essential, attributes for IR metadata, they cannot be the only sought-after characteristics. Transparency and approachability should also be objectives for IR metadata, which ultimately entail re-evaluating and reimagining the purpose of metadata documentation, its representation, and its intended audience.

**Conclusion**

The results from both the review of websites and the survey responses were statistically similar and showed that a little over half of all ARL institutional repositories made metadata documentation available via their websites. However, nearly half of the IRs do not publicly provide metadata documentation. One contributing factor may be a limited idea of who should be considered a “user” of an IR. The survey did not explicitly ask why metadata documentation was not provided, but the question merits future investigation.

This study suffered from several methodological weaknesses. It is impossible to exactly compare the website data with the survey results because there were questions unique to each method. Other potential deficiencies were survey distribution and a lack of participant eligibility verification. The solicitation for survey participation was sent to e-mail lists not limited to ARL members, since there was no list only for staff of ARL member libraries below the rank of library director or that was not specific to an ARL initiative. Finally, the survey employed some confusing language when referring to metadata documentation. The phrase “metadata documentation and/or guidance”
was employed. This caused uncertainty among some survey participants who viewed documentation and guidance as having separate meanings.

There are multiple opportunities for future research that will build up to larger questions of whether publicly available metadata documentation results in higher quality and more useful, trustworthy, and approachable metadata. To get closer to answering the research question, does providing metadata documentation publicly result in higher quality metadata, the author intends to examine metadata records from repositories with documentation versus those without to address whether documentation results in higher quality metadata.

Additionally, a more thorough investigation is needed into why IR librarians choose not to provide metadata documentation publicly. A logical next step would be to survey non-ARL academic libraries on whether they provide metadata documentation for their IRs, as well as reviewing those libraries’ IR websites for comparison to the findings in this study.

Another potential avenue would be a content analysis of institutional repository metadata documentation to determine what, if any, information is commonly provided across repositories. Such an analysis could help identify elements for standardizing this type of documentation, as well as formally differentiating between metadata policies and metadata guidelines. Finally, more research on building and maintaining community trust via metadata transparency and approachability is a timely area of inquiry.

Ultimately, the current understanding of who the appropriate audience are for institutional repository metadata documentation and what its purpose should be must be interrogated, expanded, and reframed. This exploration is desirable not only to support technical needs, such as indicating required information and data entry rules, but also to provide evidence of transparency and the continued trustworthiness of libraries to steward, preserve, describe, and propagate knowledge.

Acknowledgments

The author extends special thanks to Heidi J. Imker, Harriett Green, Joanne Kaczmarek, Timothy W. Cole, Michael Norman, Myung-Ja Han, and Micah Charles Kenfield.

Ayla Stein Kenfield is the repository services librarian and an assistant professor at the University of Illinois at Urbana-Champaign; she may be reached by e-mail at: astein@illinois.edu.
Appendix A

Institutional Repository Metadata Documentation Practices at ARL Libraries

Survey Flow

(Block: Informed consent and eligibility [3 questions])

Standard: Demographics (3 questions)
Standard: Repository management questions (4 questions)
Standard: Repository documentation questions (17 questions)

(Start of Block: Informed consent and eligibility)

Q1 Institutional Repository Metadata Documentation Practices at ARL Libraries

Investigator: Ayla Stein, metadata librarian and assistant professor in the University Library at the University of Illinois at Urbana-Champaign.

You are invited to participate in a survey for a research project that will ask respondents if and how their institutional repository’s metadata requirements are documented and made available to users. If you meet both of the following criteria, you are eligible for participation in this survey: institutional repository managers, institutional repository librarians, scholarly communications librarians or equivalent roles at their institution; those who are employed at an Association of Research Libraries (ARL) academic member library.

For the purposes of this survey, an institutional repository (IR) is an online “digital collection” that stores, preserves, makes discoverable, and provides access to the intellectual, scholarly, and research output of an institution. IRs were typically designed to hold document-based items such as conference papers and journal articles, but in many cases have expanded to host presentations, posters, research data, and other scholarly materials.

This study will take approximately 10 minutes of your time.

Your decision to participate or decline participation in this study is completely voluntary, and you have the right to terminate your participation at any time without penalty. You may skip any questions you do not wish to answer. If you do not wish to complete this survey just close your browser.

Although your participation in this research may not benefit you personally, anonymized survey results may be shared, which could help inform local institutional repository management work.

The results of this study may be published in professional journals. They may also be used for educational purposes or for professional presentations. However, no individual subject or identifying information will be shared.

There are no risks to individuals participating in this survey beyond those that exist in daily life. Your decision to participate, decline, or withdraw from participation will have effect on your current status or future relations with the University of Illinois. Participation in this project is voluntary, and the only alternative to the survey is non-participation in the project.
All reasonable efforts to keep your personal information confidential will be used, but absolute confidentiality cannot be guaranteed. When this research is discussed or published, no one will know that you were in the study. But, when required by law or university policy, identifying information may be seen or copied by:

The Institutional Review Board that approves research studies;

- The Office for Protection of Research Subjects and other university departments that oversee human subjects research;
- University and state auditors responsible for oversight of research.

When the results of the research are published or discussed in conferences, no information will be included that would reveal your identity.

If you have questions about this project, you may contact Ayla Stein via e-mail at stein@illinois.edu or by phone at 217-300-2598. If you have any questions about your rights as a participant in this study or any concerns or complaints, please contact the University of Illinois Office for the Protection of Research Subjects at 217-333-2670 or via e-mail at irb@illinois.edu. Please print a copy of this consent form for your records, if you so choose.

Q2 Please indicate if you agree or do not agree to participate in this research:

- I agree to participate in this research (1)
- I do not agree to participate in this research (2)

Q3 Does your organization provide an institutional repository? (Reminder: For the purposes of this survey, an institutional repository is an online “digital collection” that stores, preserves, makes discoverable, and provides access to the intellectual, scholarly, and research output of an institution.)

- My institution does provide an institutional repository (1)
- My institution does not provide an institutional repository (2)

Q5 Is your institution private or public?

- Private university (1)
- Public university (2)

Q7 Is your organization a(n): (please select all that apply)

- Historically black college or university (HBCU) or a predominantly black institution (PBI) (1)
- Hispanic-serving institution (HSI) (2)
- Asian-American and Pacific Islander (AAPI) serving institution (3)
- Other minority serving institution (MSI) (4)
Q8 How many students are enrolled at your institution?
- 1–9,999 (1)
- 10,000–19,999 (2)
- 20,000–29,999 (3)
- 30,000–39,999 (4)
- 40,000–49,999 (5)
- 50,000+ (6)

(End of Block: Demographics)

(Start of Block: Repository management questions)

Q10 How many individual full-time staff are dedicated to the institutional repository?
(For the purposes of this survey, a full-time employee is someone who is employed by your library for at least 35 hours per week.)
- 0–2 (1)
- 3–4 (2)
- 5+ (3)

Q11 How many individual part-time staff are dedicated to the institutional repository?
(For the purposes of this survey, a part-time employee is someone who is employed by your library for 34 hours per week or less.)
- 0–2 (1)
- 3–4 (2)
- 5+ (3)

Q12 How is your institutional repository designed to be managed? This may be different from how repository management operates de facto. Please choose the option that best fits your repository. (Note: The examples for each of the options below are intended to provide a general idea of what each option means and are not prescriptive.)
- Distributed: Each community or equivalent has a designated administrator who handles the daily management of the community, which may include (but is not limited to) the following: approving individuals to submit materials to their community; reviews and approves submitted items and/or metadata as deemed necessary; creates collections within the community; performs outreach to their community’s stakeholders, deposits materials on behalf of researchers, etc. (Note: In this instance, a community (or its equivalent) is a repository container that holds collections of repository materials or individual repository items, depending on setup of the IR. Communities may be formed at various granularities which may vary for each repository. They are often setup at the college/school/faculty or department levels) (1)
- Centralized: Administrative tasks for all communities across a repository are handled by a single top-level administrator or administrative team. These tasks may include (but are not limited to) the following: approving individuals to submit materials for all communities or collections; reviews and approves all deposits into the repository (both individual submissions and batch uploads;
creates new communities; creates new collections within existing communities; performs outreach to all types of repository stakeholders, etc. (2)

- Both: Repository is designed to have designated administrators for specific communities as well as centralized management across repository communities. (3)
- Other: Please describe how your repository is designed to be managed (4)

Q13 How are materials deposited into your institutional repository?

- Unmediated self-deposit: Any permitted user can submit individual scholarly works to the institutional repository. There is no repository-level requirement that all materials and metadata be reviewed before final acceptance into the repository. (1)
- Mediated self-deposit: Any permitted user can submit individual scholarly works to the institutional repository. There is a repository-level requirement that all materials and metadata be reviewed by repository staff before final acceptance into the repository. (2)
- Repository staff-proxy deposit: No users are permitted to submit individual scholarly works to the institutional repository; submission of all materials is handled by repository staff. Researchers must send their files to repository staff for their work to be added to the repository. (3)
- Other: Please describe how materials are deposited into your institution’s IR (4)

Q14 Does your institutional repository provide documentation and/or guidance on your institutional repository’s metadata?

- Yes, documentation/guidance is provided on my institutional repository’s metadata (1)
- No, documentation/guidance is not provided for my organization’s institutional repository metadata (2)

Q15 Does the metadata documentation/guidance provided indicate the required metadata elements?

- Documentation or guidance describing the required metadata elements is provided. (1)
- No documentation or guidance about required metadata is provided (2)
Q16 If no documentation or guidance about required metadata is provided, why not? Please select all that apply:

- Required metadata elements are indicated in submission workflow (1)
- Required metadata is pulled from file properties (2)
- Not enough staff to create documentation (3)
- Documentation is in development (4)
- Metadata creation is handled by staff (5)
- Other (6) ________________________________________________

Q17 Does the metadata documentation/guidance provided describe how to input metadata (i.e. data entry rules) when submitting an item to the institutional repository?

- Yes—the metadata input documentation/guidance is integrated with other (non-metadata) documentation (e.g. the “how to submit an item to the repository” documentation) (1)
- Yes—documentation/guidance for metadata entry is provided as standalone data entry rules and is not incorporated into other documentation (2)
- Yes—all institutional repository metadata information is provided together (3)
- Metadata input documentation/guidance is not provided (4)

Q18 If metadata input documentation is not provided, why not? Please select all that apply:

- Data entry rules documentation is on the to-do list but is not a high priority (1)
- There are no free-text metadata fields (2)
- Not enough staff (3)
- Documentation is in development (4)
- Other (5) ________________________________________________

Q19 Does the metadata documentation/guidance provided describe what metadata is automatically generated by the repository upon submission?

- Yes, there is documentation/guidance provided that describes the information automatically collected by the repository software upon submission of an item. (1)
- No, the metadata documentation/guidance available does not provide information about automatically generated metadata. (2)

Display this question:

If Does the metadata documentation/guidance provided describe what metadata is automatically generated . . . = No, the metadata documentation/guidance available does not provide information about automatically generated metadata.
Q20 Why is information on automatically generated metadata not included in the metadata documentation/guidance? Please select all that apply:

- Information on system generated metadata was not considered necessary to include in documentation (1)
- Users have not asked for it (2)
- We forgot to include this information (3)
- Other (4) ________________________________________________

Q21 How is the institutional repository metadata documentation made available to users?

- Publicly available (1)
- Available by request (2)
- Login required (3)
- The metadata documentation is for staff use only (4)
- Other (5) ________________________________________________

Q22 Has the metadata documentation/guidance ever been downloaded or clicked on by nonlibrary staff users?

- Yes, the metadata documentation/guidance has been downloaded or clicked on by nonlibrary staff users. (1)
- No one other than library staff has accessed the metadata documentation for our institutional repository. (2)
- I don’t know—we don’t track access statistics to our repository documentation. (3)

Display this question:

If Has the metadata documentation/guidance ever been downloaded or clicked on by nonlibrary staff users . . . = Yes, the metadata documentation/guidance has been downloaded or clicked on by nonlibrary staff users.

Q23 How often, on average, is the metadata documentation/guidance accessed by a nonlibrary staff user per month?

- 1–10 (1)
- 11–20 (2)
- 21–30 (3)
- 30+ (4)

Q24 Do you know how the metadata documentation/guidance is being used? Please select all that apply:

- Documentation/guidance is used prior to submission of items (1)
- Documentation/guidance is used simultaneously alongside an item submission (2)
- Documentation/guidance is used as an example for an LIS or archival course (3)
- Documentation/guidance is used as a model for documentation/guidance at another institution. (4)
- Other use of our metadata documentation: (5)
Q25 How often are repository staff contacted by depositors for help inputting metadata on average per month?
- Repository staff are never contacted for help entering metadata (1)
- 1–10 (2)
- 11–20 (3)
- 21–30 (4)
- 30+ (5)

Q26 Please select your top 3 most pressing metadata concerns for your institutional repository’s metadata:
- Spelling errors in user-provided metadata (1)
- Variations of creator and/or contributor names (2)
- Lack of authority control (3)
- Data values do not match metadata field (e.g. information is entered into the incorrect field) (4)
- Missing required metadata (5)
- Not enough descriptive information to grasp what item contains without opening file (6)
- Entered data is in all caps instead of sentence or title case (7)
- Author name parts are entered incorrectly, e.g. “first name, last name” instead of “last name, first name” (or incorrectly when based on my IR’s metadata policies) (8)
- Encoding errors when non-Roman alphabet symbols are entered (e.g. math or Greek symbols, non-Roman language alphabets, etc.) (9)
- Use of noncontrolled vocabulary terms when a metadata element specifies a required controlled vocabulary. (10)
- Inconsistent representation of metadata values other than personal names, such as serials/report series names or geographic locations (11)
- Unable to incorporate use of personal identifiers such as ORCID (Open Researcher and Contributor Identifier) ID, ResearcherID, ISNI, or Scopus ID (12)
- Other: (13) ________________________________________________

Q27 Do you think that your library is adequately staffed?
- My library has enough staff to adequately develop, maintain, assess, and improve collections and services (1)
- My library needs additional staff to fulfill all of our responsibilities (2)
- My library is adequately staffed but the institutional repository team is not (3)
- Other (4) ________________________________________________

Q28 Are you willing to provide access to your repository’s metadata documentation to the researcher?
- Yes, here is the URL: (1)
- No, I don’t want to share our metadata documentation (2)
- Yes, I will upload the documentation as an attachment (3)
Display this question:
If Are you willing to provide access to your repository’s metadata documentation to the researcher? = Yes, I will upload the documentation as an attachment

Q29 Please upload the metadata documentation or guidance for your institutional repository.
Q30 Are there any other comments you would like to share about metadata management in institutional repositories?
(End of Block: Repository documentation questions)
Appendix B

List of Reviewed Institutions

Arizona State University
Auburn University
Boston College
Boston University
Brigham Young University
Brown University
Case Western Reserve University
Colorado State University
Columbia University
Cornell University
Duke University
Emory University
Florida State University
George Washington University
Georgetown University
Georgia Tech University
Harvard University
Howard University
Indiana University
Iowa State University
Johns Hopkins University
Kent State University
Louisiana State University
Massachusetts Institute of Technology
McGill University
McMaster University
New York University
North Carolina State University
Northwestern University
Ohio State University, The
Oklahoma State University
Pennsylvania State University
Princeton University
Purdue University
Queen’s University
Rice University
Rutgers University
Southern Illinois University Carbondale
Stony Brook University, the State University of New York
Syracuse University
Texas A&M University
Texas Tech University
University of Alberta
University of Arizona
University of British Columbia
University of Calgary
University of California Berkeley
University of Chicago
University of Cincinnati
University of Colorado Boulder
University of Connecticut
University of Delaware
University of Florida
University of Georgia
University of Guelph
University of Hawaii
University of Houston
University of Illinois at Chicago
University of Illinois at Urbana-Champaign
University of Iowa
University of Kansas
University of Kentucky
University of Louisville
University of Manitoba
University of Maryland
University of Massachusetts Amherst
University of Miami
University of Michigan
University of Minnesota
University of Missouri
University of Nebraska Lincoln
University of New Mexico
University of North Carolina at Chapel Hill
University of Notre Dame
University of Oklahoma
University of Oregon
University of Ottawa
University of Pennsylvania
University of Pittsburgh
University of Rochester
University of Saskatchewan
University of South Carolina
University of Tennessee
University of Texas
University of Toronto
University of Utah
University of Virginia
University of Washington
University of Waterloo
University of Wisconsin Madison
Vanderbilt University
Virginia Polytechnic
Washington State University
Washington University in St. Louis
Wayne State University
Western University
Yale University
York University
Notes


4. Ibid.

5. Park, “Metadata Quality in Digital Repositories.”


12. Ibid.


22. Simon Fraser University in British Columbia, Canada, and Virginia Commonwealth University in Richmond became ARL members effective January 1, 2018. They were not counted for the purposes of this project.
23. Although Université Laval in Quebec, Canada, otherwise meets this criteria, the author found only French versions of its institutional repository documentation, prompting the mandate of documentation available in English.
24. Originally the University of California campuses were counted separately, but the data the author collected were the same for each institution, so they were merged into a single row.
30. Probets and Jenkins, “Documentation for Institutional Repositories.”
34. Park, “Metadata Quality in Digital Repositories.”
38. Ibid.
42. CCSDS, “Audit and Certification of Trustworthy Digital Repositories.”
43. CoreTrustSeal is a new organization created from the merger of the Data Seal of Approval and the ICSU (International Council for Science) World Data System organizations.