Understanding Metadata Needs when Migrating DAMS

Ayla Stein  
University of Illinois at Urbana-Champaign, USA  
astein@illinois.edu

Santi Thompson  
University of Houston, USA  
sathompson3@uh.edu

Abstract

This study identifies and explores metadata needs associated with migrating to a new Digital Asset Management System (DAMS). Drawing upon results from a 2014 survey, titled “Identifying Motivations for DAMS Migration: A Survey,” this paper analyzes survey questions related to metadata, interoperability, and digital preservation. Results indicate three distinct metadata needs for future system development, including support for multiple or all metadata schema, metadata reuse, and digital object identifiers. While some of these needs resemble long-standing conversations in the professional literature, others offer new areas for system development moving forward.

Keywords: metadata, digital asset management systems

1. Introduction

In the last two decades, digital asset management systems (DAMS) have become important tools for collecting, preserving, disseminating, and making discoverable digitized and born digital content to library users. During that time libraries have selected a variety of DAMS to manage their digital assets, including proprietary systems (Ex Libris’ DigiTool and OCLC’s CONTENTdm), open source platforms (Greenstone, Fedora, Islandora, and DSpace), and homegrown solutions. Over time libraries have begun re-assessing DAMS based on the changing needs of users, the expanding skill sets of librarians and staff, and the evolution of web technologies. As libraries engage in this process, some choose to migrate from one DAMS to another.

The data referenced in this paper is drawn from “Identifying Motivations for DAMS Migration: A Survey,” which identified thirteen topical categories for migrating from one digital asset management system (DAMS) to another. Researchers focused the survey on systems used to provide access to primary source research materials. The scope emphasized that the survey did not focus on systems used exclusively as institutional repositories, which the researchers define as repositories that provide access to university scholarship. This paper analyzes a subset of the responses which focus on the topics Metadata Standards, Interoperability, and Preservation. In the survey the researchers defined each of the three categories as:

- **Metadata Standards:** The “New DAMS’s” support of established metadata standards, user generated metadata, and linked data technologies.
- **Interoperability:** The “New DAMS’s” ability to export metadata into other DAMS and digital program environments. The “New DAMS” should support international and/or industry standards for interoperability, including OAI-PMH, Z39.50, and SRU/SRW protocols.
- **Preservation:** The integration of preservation strategies into the “New DAMS”, including fixity verification and the creation of checksum values, backups, synchronization, and/or the generation of archival information packages (AIPs).
The researchers believe that results from this data may give insight to the question, “What are the metadata needs for migrating from one DAMS to another?” Understanding these needs could help align future DAMS development and adoption with emerging metadata trends and initiatives.

2. Literature Review

Metadata is a core element to any library DAMS. This literature review examines works that focus on the relationship between metadata and DAMS functionality in order to compare established practices, identified gaps in the literature, and emerging needs from survey results.

Attention to relationship between metadata and library DAMS has been diverse. Some information professionals have addressed broad ways that metadata supports core DAMS functionality. Payette (1998) identified several library functions, including resource discovery, access and use, preservation and administration, and persistent identifiers. Others have focused on specific tools and features. Lagoze et al. (2005) discussed how metadata automation comes in several different flavors, including: detection of embedded metadata within ingested digital objects and auto-generated metadata values. Tools to create and manage both traditional and non-MARC metadata are another significant concern. In Zeng et al.’s 2009 survey, they found that survey respondents were concerned with a lack of metadata tools that are easy to use and do not require a steep learning curve.

User contributions is another metadata feature in library DAMS that has received attention in the professional literature. In order for DAMS to meet the needs of users, they need to upgrade to Web 2.0., which is characterized by user contributions and interactions with online content (Beal, n.d.). In their comparison survey of DAMS, Andro et. al. (2012) identified several systems that enabled users to make contributions to metadata, either through the process of “annotating” or “commenting” (p. 82). Others have focused on issues that arise from implementing user contributions in a DAMS. Lagoze et al. (2005) discuss how user contributed content creates complications for system distinctions between metadata and data. They note “...one of the useful forms of contextual information is annotations. Are these metadata (about something) or data in their own right? There is no one answer, but an architecture that imprints the distinction between data and metadata makes it difficult to deal with such ambiguities” (p. 6). Still other parts of the literature emphasize curatorial and user engagement possibilities. For example, crowd-sourced additions might augment or even replace time-intensive and expensive metadata creation and maintenance work (Mitchell and Gilbertson 2008).

Some studies have identified intersections between metadata and DAMS development that need further research and support. In their comparison of 10 DAMS, Andro et.al (2012) compared how systems supported multiple metadata schema, including non-traditional library schema. The authors discovered that all or most systems supported some degree of “library” metadata (including DC, MODS) and “archives” (EAD) (p. 80). However, less than half of the systems supported “research,” “learning,” and “photo” metadata (p. 80). This research suggests that current systems lack support around metadata schema and functions that describe research and other activities outside of the library environment. Additionally, Goh et. al. (2006) proposed that systems should support multiple metadata schema since virtually all of the systems evaluated in their study supported only core standards (such as MARC21 and Dublin Core) (p. 367). Furthermore, Han et. al (2010) suggested that future research could better configure complex objects in CONTENTdm to maximize discoverability and interoperability (p. 77).

Finally, metadata contributes to DAMS functionality through enabling interoperability. Because DC is a flexible, simple schema, it is well suited for promoting interoperability among other systems. Han et. al. (2010) argued that drawing on best practices could help promote interoperable metadata as well as eliminate metadata problems derived from inconsistencies in
localized practice (p. 74-75). Zeng et al. (2009) noted that there is also significant interest in designing systems that can natively handle or map between different metadata schemas. Additionally, Lagoze et al. (2005) wrote that “we should be wary of throwing out collections of cataloging records, and ignoring the value that uniform metadata has for ‘order making’ over heterogeneous information. However, we need to incorporate these catalog records into a richer foundation that represents...complex relationships and a host of other complexities” (p. 6).

Linked data technology is widely considered to be the solution to non-metadata centric systems (Solodovnik 2011). While interoperability issues may be somewhat ameliorated by the implementation of linked data, as it currently stands, the legacy methods of developing metadata vocabularies, in disciplinary silos, is being carried over to the Semantic Web: “a major source of interoperability problems on the Semantic Web is still due to the use of different value vocabularies supporting metadata descriptions in different linguistic communities” (Solodovnik 2011, p.10). It’s clear that the use of Linked Data technologies in and of themselves will not be enough to promote interoperability. It will require cross-disciplinary and inter-institutional collaboration. The success of the schema.org vocabulary could arguably be attributed to the fact that it was developed and implemented by the three largest search engine corporations: Google, Yahoo!, and Bing (O’Connor 2011).

The results of this survey build upon many of these themes, including the continued need for supporting multiple metadata schema, sharing data among systems in new ways, and the future role of linked data. It also begins to expand the discussion around differing development areas, including metadata reuse among users and the role of digital object identifiers in library DAMS intended to curate and make accessible digitized special collections materials.

3. Methodology

To complete this study, researchers analyzed a subset of data from a larger investigation that seeks to identify motivations for migrating from one DAMS to another. Using a survey as their instrument, they solicited responses by emailing calls for survey participation to eight listservs related to digital curation from July through September 2014. In order to qualify for the survey, respondents had to fulfill one of the following three eligibility categories:

1. Institutions had completed migration from the “Old DAMS” to the “New DAMS”
2. Institutions were currently migrating from the “Old DAMS” to the “New DAMS”
3. Institutions selected a “New DAMS” but had not started the migration process
4. If institutions selected “none of the above,” the software automatically ended the survey. Since the researchers solicited anonymous responses from listserv subscribers, they did not have the information needed to calculate a response rate. Once initiated, the survey had a completion rate of 47%. After removing ineligible entries, the researchers had 49 responses to analyze for this study. Over half of the eligible responses came from academic libraries. For more information, see Table 1: Which of the following best describes your library?

---

1The listserves included: The Code4Lib main listserv; DigLib, the International Federation of Library Associations (IFLA)'s digital library focused listserv; DigiPes, an American Library Association (ALA) listserv focused on digital preservation issues; Archives and Archivists, the main listserv for the Society of American Archivists; the Research Data Access and Preservation (Rdap) focused listserv from the Association of Information Science and Technology (ASIS&T); DLF-ANNOUNCE, a listserv from the Digital Library Federation; pasig-discuss, the discussion listserv for ASIS&T’s Preservation and Archiving Special Interest Group (PASIG); and acr-igdc-l, the listserv for the Association of College and Research Libraries' (ACRL) Digital Curation Interest Group.
TABLE 1: Which of the following best describes your library?

<table>
<thead>
<tr>
<th>Response Type</th>
<th>Total Number of Responses</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Academic Library</td>
<td>30</td>
<td>61</td>
</tr>
<tr>
<td>Research Library</td>
<td>8</td>
<td>16</td>
</tr>
<tr>
<td>Public Library</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>Special Library</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Special Collections Libraries or Archives</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Government Library</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Other</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Museum Library</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

To create the survey, the researchers crafted specific questions around thirteen topics related to DAMS evaluation, including:

- Implementation & Day-to-Day Costs
- User Administration
- Organizational Viability
- Technical Support
- System Administration
- Extensibility
- Information Retrieval & Access
- Content Management
- Preservation
- User Interface Customization
- Interoperability
- Reputation
- Metadata Standards

Survey questions for these topics were designed to be either a Likert scale of 1 [Not Important] to 4 [Very Important] or select all that apply. The survey asked for key demographic information to help the researchers understand how institutions prioritized potential motivations. Demographic questions required respondents to select and/or self-identify the “Old DAMS” and the “New DAMS.” Next, the survey asked respondents to choose the top five motivations from one of the thirteen topics and then prioritize those five selections in order of importance. At that point, respondents answered questions from the five topics they identified.

Since the scope of this paper is to understand the relationship between metadata needs and DAMS migration, the researchers identified questions that addressed metadata features and functionality. Researchers used the survey reports feature in Qualtrics to generate descriptive statistics for the selected questions, including total amount, statistical mean, and standard deviation. They drew upon these reports to formulate conclusions and identify future research areas.²

² The Qualtrics reports also included minimum and maximum values, as well as variance.
4. Results

Analyzing the data\(^3\), researchers determined whether certain metadata features were important or not important to respondents.

<table>
<thead>
<tr>
<th>Survey Question</th>
<th>Survey Answer</th>
<th>Total Number of Responses</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>What descriptive metadata standards/schema did you desire the &quot;New DAMS&quot; support?</td>
<td>Dublin Core</td>
<td>19</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>MODS</td>
<td>16</td>
<td>76</td>
</tr>
<tr>
<td></td>
<td>EAD</td>
<td>12</td>
<td>57</td>
</tr>
<tr>
<td></td>
<td>MARC</td>
<td>10</td>
<td>48</td>
</tr>
</tbody>
</table>

As can be seen in Table 2, researchers considered results that registered mean responses higher than 3.0 and a standard deviation of less than 1.0 to be important considerations for institutions migrating to a new DAMS. These included:

- “The ability to support multiple metadata schema”
- “The ability to support local metadata standards and practices”
- “The new DAMS supports digital object identifiers”
- “The ability to export all or part of the metadata for reuse”
- “The ability to allow other digital library environments to harvest its content”

Alternatively, researchers considered results that registered mean responses lower than 3.0 and/or a standard deviation at or above 1.0 to be less important considerations for institutions migrating to a new DAMS. These included:

- “The new DAMS supports linked data technologies”
- “The ability to support user generated metadata such as tags or folksonomies”
- “The new DAMS automates metadata creation”
- “The new DAMS supports personal digital identifiers”

Other responses demonstrate the diverse needs that future DAMS should address to remain relevant to the cultural heritage community.

\(^3\) Researchers are actively working with the data from this survey to complete another manuscript for publication. However, data are available upon request to the authors. Once published, the researchers will make the data from this project freely accessible via a repository.
While Dublin Core was the most popular response for descriptive metadata, several other standards/schema also had a high number of responses, which suggests that future systems should support multiple descriptive schemas. Additionally, the researchers received several free text responses that said DAMS should support all metadata schemas or should be schema-less. All respondents desired technical metadata to be automatically created by the DAMS. A majority of participants also expected that preservation metadata would be collected systematically. Future systems should support and generate METS records, as well as document PREMIS events as part of their core functionality. For interoperability, respondents favored using OAI-PMH and APIs over other methods to share metadata with other systems. In regards to linked data, RDF/XML and JSON are the most popular serialization formats for expressing metadata as linked data. Concerning identifiers, DOIs appear to be the most widely needed object identifiers for future systems. Additionally, if systems choose to support personal digital identifiers (PDIs), they...
should particularly consider ORCID, as well as authority identifiers such as ISNI, and authority schemas like the Metadata Authority Description Schema (MADS).

5. Discussion

Researchers drew upon response data from several survey questions to answer the research question: “What are the metadata needs for migrating from one DAMS to another?” Creating systems that support all or multiple types of metadata schema was one important need derived from survey results. The responses to the survey question “The ability to support multiple metadata schema” showed respondents desired more metadata flexibility from DAMS. The follow up questions “What descriptive metadata standards/schema did you desire the "New DAMS" to support?” and “What administrative, preservation, structural, and/or technical metadata standards did you desire the "New DAMS" to support?” affirm metadata practices that are commonly used today among institutions. For example, Dublin Core, METS, and PREMIS remain the most popular schema overall for a New DAMS to support. Since the survey did not ask respondents to explain their preferences, researchers could only speculate as to why those completing the survey selected these specific schemas. Survey results also showed that a majority of respondents desired support for other metadata schema. In addition to desiring support for DC, respondents also favored MODS, and EAD for descriptive metadata, while a still sizable number also preferred MARC and VRA Core.

Combining these results with the favorable support of another survey question, “The ability to support local metadata standards and practices,” suggests a need for future systems to support multiple or all schema (either locally-derived or based on formal standards) as Goh et. al. (2010) argued (p. 367).

Another need that emerged from the survey results focused on facilitating library metadata reuse by both systems and users. The responses to the survey question related to “The ability to allow other digital library environments to harvest its content” suggested that respondents still highly valued the ability to make their data interoperable with other library DAMS. A follow up question, “What interoperability methods and/or standards did you desire the "New DAMS" to support?” showed that OAI-PMH remains the most popular aggregation method for respondents, surprising researchers who thought the growing system development around APIs would have made it the most popular method. Despite libraries growing comfort in the technological realm, implementing new technologies such as APIs still requires specialized knowledge and skills, which may be why established protocols such as OAI-PMH are still in high demand. There may also be a desire to support technologies developed within the library domain and some inherent resistance to external innovations. With limited resources and time, librarians may prefer to stay with the technologies they have helped to create and support over time.

Complementing system reuse, results from the survey question “The "New DAMS" has the ability to export all or part of the metadata for reuse” showed how respondents favored system functionality around user reuse. Often metadata records contain rich contextual information about digital objects that, in itself, can be valuable data for research. Because the amount of attention focused on reusing data, from data sets to metadata in digital humanities projects, has increased over the last several years, the researchers were not surprised by this need. Since most of the literature dedicated to selecting DAMS and to the role that metadata plays in DAMS functionality do not address user reuse of metadata, the researchers believe that a gap exists in the literature around designing DAMS for metadata reuse by the user; this a gap should be addressed in future research.

A third need focused on future DAMS supporting digital object identifiers. Results from a survey question that explored “The new dams supports digital object identifiers” suggested that

---

4 Schema focused on particular formats or content types (VRA Core and PB Core, for example) were not as highly selected; if it is not possible for system to support all schema, it is unclear just how integrated future systems should be with these schema.
respondents desired a future system that has the capability to generate identifiers for digital objects. A follow up question, “What digital object identifiers did you want the "New DAMS" to support?” showed that respondents favored Digital Object Identifier System identifiers (DOIs) specifically, which surprised the researchers because of the cost implications related to DOIs, as well as the lack of anecdotal evidence of libraries adopting DOIs for their digitized collections. Readers should note that there are some limitations around the results of this particular question based on an error in the survey instrument. Researchers included ezid as possible response for the follow up question related to digital object identifiers. Since ezid mints identifiers (dois and ARKs), it should not appear in the question. Additionally, the scope of the survey was based on DAMS intended to curate digitized special collections content. However, some institutions may have one unified DAMS that fulfills multiple purposes, including disciplinary or institutionally-based repositories, which have a wider adoption of digital object identifiers. In any case, future research on the role of digital object identifiers in digital library/digital collections environments should be explored further.

While three needs emerged from the survey data, the researchers concluded that the response data to the other topics related to metadata and future DAMS development could not be applied to the research question because these are areas that require more in-depth research and investigation.

The responses to the survey question “The New DAMS supports linked data technologies” indicated a lack of consensus on whether or not linked data technologies were considered necessary for New DAMS. While fifteen respondents indicated that support of linked data technologies were considered ‘important or very important’, seven respondents indicated ‘not important’ or ‘somewhat important’. The lack of consensus reflects the present status of applied linked data technologies. Until relatively recently, linked data was, and still often is, an abstract or intangible concept. While research, investigation, and infrastructure development on library linked data has been underway for several years (Baker et. al. 2005; Library of Congress, n.d.), it was not until the release of Fedora 4 (DURASPACE 2014), and to a lesser extent Kuali OLE (Kuali n.d.), that native linked data library systems became readily available. Even between these two systems, only Fedora 4 could function as a DAMS. There is still a significant amount of work that needs to be accomplished before linked data technology is within reach of most libraries.

The responses to the survey question “The ability to support user-created metadata such as tags or folksonomies” also indicated a lack of consensus. Responses were almost evenly distributed, with ten respondents indicating it was ‘Not important or somewhat important’, and twelve indicating ‘important or very important’. These results were somewhat surprising in light of the significant interest and optimism regarding user-created tags in the literature (Lagoze et. al. 2005; Mitchell and Gilbertson 2008). The formation of questions may have also impacted results. The researchers focused entirely on user-created vocabularies, and did not include examples of added-value metadata, e.g. annotations. The researchers suspect that the type of user-created metadata needed in DAMS has changed over time, (especially with the proliferation of tablets, “phablets” (Oxford English Dictionary 2015), and touchscreens) and research-oriented user-metadata features, like highlighting and annotating, would be rated more highly. This topic is an area of future investigation that the researchers hope to explore further with institutional, data, and scholarly repositories.

Responses to the question “The New DAMS automates metadata creation” indicate that participants do not consider automated metadata creation to be a required function of the New DAMS. These results were surprising to the researchers given the attention that the literature paid to the varieties of metadata automation (Lagoze et. al. 2005). The researchers believe that the results are partially due to poor wording of the question, which does not reflect multiple types of metadata automation. A follow-up question, “What metadata did you desire the "New DAMS" to automatically create?” asked respondents to select-all-that-apply with possible responses of ‘technical metadata’, preservation metadata’, or ‘other/free text’. This question did not clarify
what researchers meant by automated metadata creation. A more appropriate question to ask would have focused on specific use cases for automated metadata creation.

Responses to the question “The New DAMS supports personal digital identifiers” conclusively indicate that personal digital identifiers (PDIs) are not necessary for New DAMS to support. This result did not surprise the researchers given that PDIs such as ORCID and ResearcherID are far more prevalent in institutional and scholarly DAMS than those focused on digital library collections. The connection between DAMS and PDIs is an area of future inquiry for the researchers.

Researchers have identified several limitations with the composition of the survey and the results derived from it. Because there is no definitive DAMS registry encompassing all libraries, the researchers cannot determine whether or not the results are statistically significant. Furthermore, the data are not necessarily based on a representative or random sample. Since researchers relied on voluntary participation from those who subscribed to certain listservs, they have no way of knowing the total number of possible participants or calculating a response rate.

6. Conclusion

The purpose of this investigation was to understand metadata needs when migrating from one DAMS to another. After analyzing both the existing literature and the survey results, the researchers have identified three specific needs:

1. Support for multiple or all metadata schema
2. Support for metadata reuse among other library DAMS as well as among users
3. Support for digital object identifiers

Viewed as metadata use cases for future DAMS developers, including both open source and proprietary, these three needs indicate that future DAMS should continue to embrace flexibility in metadata creation, management, export, and interoperability. In some ways, they mirror long-standing conversations in the professional literature. The desire to accommodate multiple schema and share it with a variety of library systems are not new or under-researched areas within the library profession; however, these results do suggest that librarians and system developers have yet to bridge critical functionality gaps. To address these needs, conversations around metadata should be occurring from the earliest stages of system planning and development. Likewise, metadata specialists should be involved at all stages, from design to migration. Combining these needs with the desire to expand metadata reuse for users and to generate DOIs for rare and unique digitized materials offers a variety of development areas moving forward.

References


However, several free text responses indicated the desire for systems to support authority files, including universal authority files, such as ISNI, and locally developed authorities through the support of MADS.


