October 2012

Using Data Profiles to select Digital Asset Management Systems

Lisa Zilinski
University of South Florida, ldz@andrew.cmu.edu

S. Lorenz
University of South Florida in Lakeland

Follow this and additional works at: http://repository.cmu.edu/lib_science

Part of the Library and Information Science Commons

Published In
Proceedings of the American Society for Information Science and Technology.
ABSTRACT
Many grants, including grants from the National Science Foundation (NSF) require recipients to utilize data management plans; some grants, like those given by NSF, require recipients to place their data in searchable data repositories. It is difficult to find one data repository that meets every researcher’s needs. To solve this problem, we utilized Purdue’s Data Curation Profile Toolkit to interview researchers at our university and assist in developing data profiles/plans. These profiles provided the Digital Asset Management (DAM) project team with valuable information about the researchers’ data needs. The DAM project team created a matrix of the researchers’ data needs and used the matrix to evaluate the credentials of 24 Digital Asset Management Systems (DAMS). The team was able to select the DAMS that met the needs of researchers. Using this method, any university could select a DAMS and be confident that it would meet the needs of the researchers it was intended to serve.

KEYWORDS
Data Management, Data Curation, Digital Asset Management, Data Profiles, Data Management Plans

INTRODUCTION
Data curation is complex. Researchers have reservations regarding data curation because of the perceived threat of intellectual property theft. Many researchers are also concerned about the usefulness of raw data being placed into a repository: if the raw data is undecipherable to others, what good can come from making it publicly accessible? Institutions have concerns over the cost of data curation and librarians see data curation as a gargantuan task with no reasonable path or solution (Purdue, n.d.).

In 2011, the National Science Foundation (NSF) brought the problem of data management to the forefront by requiring researchers to include data management plans in their grant proposals and by requiring researchers to share their data, generally via a publically accessible repository (NSF, 2011). Regardless of NSF requirements, data management is a problem that needs to be solved. Higgins notes that digital curation is no longer bound to only use preservation techniques that require data to be stored in a vault with only one or two authorized users (2011). To further her point, Digital Asset Management Systems (DAMS) offer many efficient solutions that incorporate metadata and in so doing, DAMS help to increase access for all users.

According to McCord (2002), “a DAMS infrastructure can ingest digital assets, store and index assets for easy searching, retrieve assets for use in many environments, and manage the rights associated with those assets,” and is “used to capture, catalog, store, and manage digital assets” (p. 1). These assets can range from assets that are born digital and those that are digitized (Currall & Moss, 2009).

The question is: how can libraries select a DAMS that will meet the needs of their researchers, universities and funding institutions while providing sustainable digital preservation and access?

METHOD
The Digital Asset Management (DAM) project team completed two sections for this research: data curation profile development and digital asset management system (DAMS) selection.

The method for developing the data profiles consisted of several steps. The first step was to identify researchers who were in need of developing data management plans. Four current research projects on the university campus were identified. The research leaders of each project were contacted and asked to participate in the development of a data curation profile. Interviews were conducted using the Purdue University Data Curation Profile Toolkit (Purdue, n.d.) interview template. The interview served to facilitate communication between the library and the researchers. All interviews were recorded and used for reference to build the data curation profile. The interviews allowed the researchers to openly and thoroughly discuss their data, as well as provided opportunities for the DAM team to ask...
questions to further identify the researchers’ data needs and requirements.

The data profile begins with a summary of the data and is followed by two main sections. The first section is comprised of details of the dataset. This section includes two subsets: (1) Overview of the research, and (2) Data kinds and stages. The second section is comprised of details regarding how the dataset is currently managed. This section includes ten subsets: (1) Intellectual property, (2) Organization and description of the data, (3) Ingest/Transfer, (4) Sharing and access, (5) Discovery, (6) Tools, (7) Linking/Interoperability, (8) Measuring impact, (9) Data management, and (10) Preservation. In these sections, Purdue’s Data Curation Profile identified the unique needs of each researcher.

The method for selecting a DAMS consisted of several steps. The data collected from the data curation profiles were examined to determine the following: Who were the end users? Why types of files/objects need to be stored? Who provides the files/objects? How are they provided? How are they organized? How do they need to be accessed? How long do they need to be embargoed or stored? The answers to these questions were compiled and broken down into a needs document. The facets were further disseminated and placed into a matrix using MS Excel (Table 1). 24 digital asset management systems were further disseminated and placed into a matrix using MS Excel (Table 1). The matrix was then further analyzed for each facet in the requirements document and recorded in the matrix. The matrix was then further analyzed for systems which provided the best fit for the researchers’ requirements. The DAM project team narrowed the field to 4 systems and further analyzed the costs, systems requirements, start-up time and feasibility.

<table>
<thead>
<tr>
<th>Solution Capabilities</th>
<th>Solution 1</th>
<th>Solution 2</th>
<th>Solution 3</th>
<th>Solution 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meta-Data</td>
<td>DC and Custom</td>
<td>Y</td>
<td>DC and Custom</td>
<td>Y</td>
</tr>
<tr>
<td>File Types</td>
<td>Any</td>
<td>Any</td>
<td>Most</td>
<td>Any</td>
</tr>
<tr>
<td>Streaming Video</td>
<td>3rd party</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Structured Authoring</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>Personalized Dashboard</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Customized Permissioning</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>Workflow Management</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Microsites/Portals</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>RSS Feed Generation</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Mobile</td>
<td>Y</td>
<td>in development</td>
<td>N</td>
<td>3rd party</td>
</tr>
<tr>
<td>Modular/Scalable</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Web 2.0</td>
<td>Most</td>
<td>Y</td>
<td>Some</td>
<td>Y</td>
</tr>
<tr>
<td>Multi-Language Support</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
</tbody>
</table>

Table 1. DAM Solutions Matrix

RESULTS

The preliminary results have lead the project team to further investigate two open-source DAMS that would best fit the university’s researchers. The next steps will include vender contact, submitting request for proposals (RFP), creating time limits, and developing budgets. The DAM project team will investigate system requirements and discuss in-house or hosted solutions.

DISCUSSION

This method is useful as it can help discern the data management needs of researchers at any institution. Rather than expect researchers to use a system that may not meet their specific needs, this method involves the researchers and reviews their needs prior to selecting a digital asset management system. In our interviews, we found the researchers were both happy to contribute and pleased that their needs were considered in the DAMS selection process.

This method also has several ancillary benefits. According to Witt et al. (2009), the profiles help to increase communication between researchers and data curators. With this increase in communication, data profiles can be used as a vehicle for helping the researchers develop controlled vocabulary. If the data profile is paired with the raw data in a repository, the data profile will give the raw data context, relevance and increased searchability. Researchers outside of the current project will be able to find and use the data in their own research rather than duplicate previous efforts (Heidorn, 2011). Data profiles can help researchers develop data management plans to meet grant requirements. Utilizing this process to select a DAMS would not only aid in managing and preserving material for the long term, but also meet the unique needs of the institution’s researchers.

There are some limitations to this method. One perceived limitation is that researchers may not be willing to participate. To the contrary, the DAM project team found that researchers at this institution responded with enthusiasm. They appreciated being consulted and were optimistic about the library’s ability to assist in developing and maintaining their data management needs. The communication facilitated by the data profile interview uncovered the opportunity for the DAMS project team and one research team to include the cost of the initial build and install of the DAMS in current grant funding. We would like to further investigate other ways to include the users in the selection process.

Time can be a significant limitation. Each interview takes one to one and a half hours (Witt et al., 2009) and creating the data profile takes an average of four to seven hours. In order to understand the needs of researchers at an institution, multiple researchers must be interviewed. Considering that each profile takes six to ten hours to complete, time is the most difficult limitation to deal with. However, the ancillary benefits of data profiles translate...
into long term utility and make the time investment worthwhile.

This poster represents preliminary findings. As the DAM project group continues research in this area, we would like to investigate the gaps in communication between researchers and data curators. While we did find that researchers were enthusiastic and positive about the library’s ability to assist them with their data management needs, it is also apparent that researchers’ data needs in this area have been overlooked in the past. Exploring this area will help libraries bridge the gap between researchers and data curators.

CONCLUSION
The goal of digital curation is to manage and preserve material in a way that ensures accessibility over the long term (Higgins, 2007; Abbott, 2008). If libraries are able to choose effective DAMS that support the needs and requirements of the institution’s researchers, then they will be able to start serving their patrons as an institutional repository for scholarly research. Thus, they will begin resolving the problem of data management.

AVAILABILITY
Purdue’s Data Curation Toolkit is freely available online at http://datacurationprofiles.org/

REFERENCES


