Considering Dynamic, Non-Textual Content when Migrating Digital Asset Management Systems

Ayla Stein; University of Illinois at Urbana-Champaign; Urbana, Illinois USA
Santi Thompson; University of Houston; Houston, Texas USA

Abstract

The researchers investigate how important managing and providing access to dynamic, non-textual media was to institutions that were migrating from one digital asset management system to another. Drawing upon data from a larger survey, the researchers used responses to categorical questions addressing extensibility, content management, and metadata to answer two queries: "Do survey results suggest that institutions are considering dynamic non-textual content when migrating to a new DAMS?" and "What trends or features (developed or added through migrating), identified from survey results, could provide solutions to this problem?" The researchers conclude that while there is an interest in managing and providing access to dynamic content, there seems to be a disconnect between respondent priorities, desired functionality, and system capabilities.

Motivation

As institutions continue to provide access to digitized and born digital collections and data, some have elected to migrate from one digital asset management system (DAMS) to another for a variety of reasons. The data referenced in this paper is drawn from "Identifying Motivations for DAMS: Migration: A Survey," which identified 13 topical categories for migrating from one digital asset management system (DAMS) to another. This paper analyzes a subset of the responses which focus on topics associated with managing and providing access to dynamic, non-textual media. With the proliferation of digitized and born digital, non-textual, dynamic content, it is necessary to design systems and services to meet this growing demand. The researchers believe that results from this study may give insight into whether or not libraries are considering implementing systems with improved access to non-static content or features that enable interaction with content in new ways.

Problem

Some information professionals have suggested that as digital objects become more complex and dynamic, and as the research conducted with digital content evolves over time, existing DAMS have struggled to keep pace, especially in terms of dynamic content access, storage, and system scalability [1][2][3]. A number of institutions have begun addressing this issue. Indiana University (IU), one of the earliest institutions to document their process, migrated from Variations to Variations 2, a home-grown DAMS to another. IU elected to migrate to another home-grown solution because they had the capability to improve access to music and video content by extending existing software functionality [2]. The Variations projects ultimately resulted in the Avalon Media System Project [4]. Preservation repositories are also increasingly storing dynamic non-textual content, such as geo-spatial data and maps [5]. The literature also discusses user

tags and annotations, which, while primarily textual at first glance, require an entirely separate metadata model than traditional textual content [6].

The authors will be responding to the following questions:

Question1: Do survey results suggest that institutions are considering dynamic, non-textual content when migrating to a new DAMS?

Question 2: What trends or features (developed or added through migrating), identified from survey results, could provide solutions to this problem?

Definitions

Researchers draw upon definitions developed by the California Digital Library's Glossary to frame the differences between simple and complex objects. Additionally, they have created a definition for Dynamic, Non-Textual data.

- Simple digital objects - "Comprised of a single content file (and its format variants or derivatives) and the metadata for that file. For example, a TIFF of the Mona Lisa, a user JPEG, a reference GIF, and the appropriate metadata would comprise a simple digital object" [7].
- Complex digital objects - "Includes two or more content files (and their format variants or derivatives) and corresponding metadata. The content files are related as parts of a whole and are sequenced logically, such as pages. For example, a complex digital object could consist of a multi-page diary scanned as TIFF images, from which are generated display images (JPEGs and GIFs), plus a transcription of the diary and the metadata for each file" [7].
- Dynamic, non-textual content – For the purposes of this study, defined to be sound, film, digital art, and complex digital objects.

Approach

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, the researchers solicited participation from eligible institutions 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.

After removing entries that were not eligible or were not fully completed, the researchers had 49 responses to analyze for this

study. Over half of the eligible responses came from academic libraries.

Table 1: Which of the following best describes your library?

Response Type	N	%
Academic Library	30	61
Research Library	8	16
Public Library	4	8
Special Library	2	4
Special Collections Libraries or Archives	2	4
Government Library	2	4
Other	1	0
Museum Library	0	0

To create the survey, the researchers crafted specific questions around 13 topics [8] 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, and it also tracked that movement when institutions migrated from the "Old DAMS" to the "New DAMS." Demographic questions required respondents to select and/or self-identify the "Old DAMS" and "New DAMS." Additionally, the survey asked respondents to choose the top five motivations from one of the 13 topics listed above and then prioritize those five selections in order from most important to least important. Respondents would then answer questions in the survey that focused on the five topics they identified. Because the survey had over 100 questions, the researchers used this method to reduce the overall time needed to complete it.

Since the scope of this paper is to understand the relationship between dynamic content and DAMS migration, the researchers identified questions from the survey that addressed topics related to acquiring, describing, preserving and making accessible dynamic, non-textual media. The questions most related to these areas fell in the following survey topics:

- **Extensibility:** The ability to incorporate additional functionality and capabilities to the "New DAMS's" via viewing and manipulating the system code base, APIs, social media integration, or other measures.

- **Content Management:** Collection content and administration in the "New DAMS", including file formats, ingest issues, scalability, and rights management information
- **Metadata:** The "New DAMS's" support of established metadata standards, user generated metadata, and linked data technologies

Researchers used the survey reports feature in Qualtrics to generate descriptive statistics for the selected questions and drew upon these reports to formulate conclusions and identify future research areas.

Results

Researchers analyzed the number of responses, mean of the responses, and standard deviation from survey questions to better understand the role that dynamic media plays when migrating from one DAMS to another. The date in the following tables aggregate responses to questions in the Extensibility, Content Management, and Metadata topic areas, respectively.

Researchers considered statements 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. This means that all questions identified in this paper, excluding "The size limit for ingesting content" were important considerations for survey respondents.

Table 2: Survey questions related to extensibility, content management, and metadata

Question	N	Mean	SD
Institutions can create their own modules/plugins/widgets/etc for the "New DAMS".	21	3.67	.58
The "New DAMS" has an available API.	22	3.64	.58
The ability to support various file formats	28	3.61	.79
The code base is available for anyone to see and use (open source).	22	3.55	.60
The capacity to display both simple and complex digital objects	29	3.48	.78
The storage capacity to hold both simple and complex digital objects	28	3.46	.79
The "New DAMS" supports digital object identifiers.	22	3.23	.97
The governing organization creates modules/plugins/ widgets/etc. to fit the needs of your institution.	22	3.00	.76
The size limit for ingesting content	28	2.86	1.01

The following responses are to questions that shed additional light into the types of material and file formats as well as the kinds of administrative and technical metadata that a new DAMS should support.

Table 3: What types of objects did you desire the system to display?

Answer	N	%
Manuscripts	24	83
Images	24	83
Maps	23	79
Books	22	76
Video Content	21	72
Audio Content	21	72
Spreadsheets	11	40
Graphs	9	31
3D Objects	8	28
Other	5	17
Research data	2	7
GIS	1	3

Audio and video content are, by far, the most popular dynamic, non-textual content in which libraries are interested in DAMS supporting. 72% of respondents designated that DAMS should display audio and video content, suggesting that future systems should render some dynamic content just as well as it renders non-dynamic types of content, including maps, books, manuscripts, and static images.

Table 4: What file formats did you desire the "New DAMS" to support?

Answer	N	%
PDF	28	97
JPEG	26	90
MP3	22	76
JPEG2000	21	72
TIFF	21	72
MP4	19	66
MOV	17	59
CSV	16	55
DOC	13	49
DOCX	12	41
KML	2	7
WAV	2	7
GIS	1	3
KMZ	1	3
ArcGIS files	1	3

Of all the file formats relating to dynamic, non-textual content, those relating to audiovisual formats were the most popular. Desired support for specific audio and video formats varied. The most desired audio format was MP3, with 76% of responses, followed by 7% of responses for WAV files. MP4 video files tallied 66% while MOV files received 59%.

Table 5: What administrative, preservation, structural, and/or technical metadata standards[1] did you desire the "New DAMS" to support?

Response	N	%
METS	18	90
PREMIS	15	75
TEI	8	40
VRA Core	5	25
MIX	2	10
PB Core	2	10
Other	0	0

The most widely used administrative, preservation, structural, and/or technical metadata standards desired were METS metadata, at 90% of responses, followed by PREMIS preservation metadata with 75%, and TEI with 40%. There was very little interest in administrative, structural and/or technical metadata beyond these.

Discussion

Question 1: Do survey results suggest that institutions are considering dynamic, non-textual content when migrating to a new DAMS?

Analyzing survey results that address desired future support for specific content types shows that institutions are inconsistent when considering dynamic, non-textual content as a motivation for migrating to a new DAMS. Results favored the DAMS supporting a diverse number of file formats (going beyond static images to include audio, moving image, GIS, and digital maps). They also signaled that institutions have an interest in managing and making available audio and video content. Researchers believe that these findings demonstrate that more institutions have an expectation to make audio and video content accessible. This development is not necessarily surprising given the increase in audio/visual holdings among institutions in addition to the expanding abilities of DAMS to curate and make accessible audio/visual file formats.

However, providing access to other dynamic content types was not as significant a concern when choosing a New DAMS. For example, there was no clear consensus on whether providing access to scholarly output or research data was an important factor when migrating DAMS. There are most likely several plausible explanations for this; content producers may deposit data into discipline base, institutional based, or data based repositories, which is a future area of research for the researchers. Managing 3D objects in a repository did not appear as important to respondents. High costs and a lack of accessible digitization equipment for 3D materials could play a role in this result. In addition to these explanations, readers should also note that the researchers' survey may hold key limitations when it comes to

---

[1]   Researchers also included questions about descriptive metadata in the survey. Responses to these questions showed an overwhelming need (all respondents saying it was either "very important" or "important") for descriptive metadata support in a new DAMS. Believing these results reflected obvious conclusions, the researchers elected not to include this data in this paper.

understanding institutional needs around scholarly and research data and 3D objects. These and other limitations will be discussed at the end of this paper.

Question 2: What trends or features (developed or added through migrating) identified from survey results could provide solutions to this problem?

Two related trends emerged from the survey results. Institutions desire the ability to expand the kinds of access and interfaces available for complex digital objects, which include dynamic, non-textual media. This trend is driven by favorable responses to the support of digital object identifiers and application program interfaces (APIs). Digital object identifiers offer unique identification for every object in a repository; combined with uniform resource identifiers (URIs), digital object identifiers become the building blocks for establishing persistent, long-term access references. The researchers also hypothesize that this interest can establish the foundation for a more interactive scholarly environment within the DAMS. APIs are defined as "a set of routines, protocols, and tools for building software applications" [9]. APIs enable institutions to generate alternative "views" and functionality, such as visualizations, timelines, and geo-referenced maps, with digital objects managed by DAMS. Expanding these areas provide institutions with new ways to serve content and functionality to users seeking dynamic, non-textual content.

Complementing expanding access and interfaces, institutions desire the ability to expand system capabilities using their own resources and developers. Results show that institutions prefer the ability to create their own tools compared to relying on the system's governing body to do so. Additionally, results favored having access to a code base that is available for anyone to open and use. This offers institutions the opportunity to develop systems that handle complex digital objects and meet the research needs of their user communities in more efficient and flexible ways.

Inconclusive Results: DAMS functionalities that support dynamic non-textual content and complex digital objects.

The researchers feel that the response data for the following questions cannot be readily applied to either of the questions asked at the beginning of the paper. These are areas that will require future investigation and research.

The survey question "The size limit for ingesting content", was the only question examined in this paper that was not considered important to the survey respondents. Out of 28 survey respondents, the mean was 2.86 with a standard deviation of 1.01. While most respondents, 18/28 indicated 'the size limit for ingesting content' was important or very important, the remaining 10 respondents indicated it was not important (3/28) or somewhat important (7/28), leaving no real consensus on the matter. The researchers found this lack of consensus curious, considering that dynamic non-textual content files such as audiovisual, digital art, and 3D objects, are often larger and more difficult to send over a network than static PDFs or image files. However, the lack of consensus does seem to align with the results from our first question, that institutions are inconsistent when considering dynamic, non-textual content in DAMS migration.

The second survey question the researchers felt was in need of further research was, "What administrative, preservation, structural, and/or technical metadata standards did you desire the

"New DAMS" to support"? Only two respondents indicated interest in PBCore, and only five indicated interest in VRA Core. The researchers expected that if institutions were considering dynamic non-textual content in their DAMS migrations, then DAMS support for PBCore and VRA Core would be much higher. One possibility for these results may be that the generic output of PREMIS, METS, and descriptive metadata schemes could be seen as sufficient by most institutions, so the granularity of VRA Core and PBCore may not be seen as necessary.

Limitations

The survey used for this study was not designed to address dynamic non-textual content in DAMS migration exclusively. Consequently, the researchers identified several limitations. The design of the survey led to an inconsistent number of responses for the 13 topics related to DAMS evaluation. When writing survey questions, the researchers did not define some terms that could be interpreted in several ways, including 'simple digital objects', 'complex digital objects', or 'open source software'. Also, the wording for several survey questions could have been more carefully chosen. For example, while the researchers inquired about 3D content in the question asking, 'What types of object did you desire the system to display?', they did not include any possible 3D object file formats in the question 'What file formats did you desire the "New DAMS" to support?'. Additionally, the researchers did not leave a free-text option for 'other' in the 'Select all the file formats that apply' question. The survey scope specified that respondents focus on motivations for DAMS as they apply to digital library or digital collections environments, not repositories designed for institutional, scholarly, or research data, which might contain dynamic, non-textual content. However, migration of institutional, scholarly, or research data repositories is an area of future investigation for the researchers. The final limitation that the researchers identified was the Anglo-centrality of the listservs to which the survey was sent. While most of the listservs that the survey was distributed to boast international subscribers, they were not necessarily focused more globally or in non-English languages. In the future, the authors will make an effort to send survey invitations to IFLA and other international listservs as well.

Conclusion

The researchers believe that when migrating DAMS there is an interest in managing and providing access to dynamic non-textual content in an environment that allows for more sophisticated user interaction and research, particularly around audio and video. However, other emerging, more complex content, including 3D objects, geospatial data, and research data, remain daunting for libraries to manage, make accessible, and preserve. Future research in the areas of institutional and data repositories, digital research environments, and more complex data types can begin to acclimate librarians with these emerging challenges.

References

[1] C. L. Borgman, "The Digital Future is Now: A Call to Action for the Humanities," Digital Humanities Quarterly, 3, 4 (2009). Retrieved March 27, 2015, from http://www.digitalhumanities.org/dhq/vol/3/4/000077/000077.html.

[2] J. Dunn, M. Notess, and R. Scherle, "Flipping the Switch: Lessons Learned from a Major Digital Library Migration Project," Presented at the Digital Library Federation Fall 2005 Forum, Charlottesville, Virginia, November 7-9, 2005.

[3] K. Masood, & A. Neatrour, "Digital Asset Management System Options: Report of the University of Utah Libraries DAM Review Task Force," MWDL Winter Webinar Series (2014) Retrieved March 27, 2015, from http://www.mwdl.org/events/DAMS_options.php.

[4] Avalon Media System Project, "The Project", (2014). Retrieved December 7, 2014, from http://www.avalonmediasystem.org/project.

[5] J. Hutař, "Archives New Zealand Migration from Fedora Commons to the Rosetta Digital Preservation System," iPRES2013 Proceedings (2013).

[6] C. Lagoze, D.B Kraft, S. Payette,& S. Jesuroga, "What Is a Digital Library Anyway?" D-Lib Magazine, 11, 1–22 (2005) doi:10.1045/november2005-lagoze.

[7] California Digital Library, "Glossary" (2014) Retrieved March 30, 2015 from http://www.cdlib.org/gateways/technology/glossary.html.

[8] Marill, J. L., & Luczak, E. C. (2009). Evaluation of digital repository software at the National Library of Medicine. D-Lib Magazine, 15(5-6). doi:10.1045/may2009-marill.

[9] Webopedia "API - Application Program Interface" (2015) Retrieved March 30, 2015 from http://www.webopedia.com/TERM/A/API.html

Author Biography

Ayla Stein is currently a Metadata Librarian at the University of Illinois Urbana-Champaign Main Library. She holds a BA in East Asian Studies from the University of Arizona and a MS of Information from the University of Michigan.

Santi Thompson is currently the Head of Digital Repository Services at the University of Houston (UH) Libraries. He holds a MLIS and MA, Public History from the University of South Carolina.