University of Nebraska - Lincoln
DigitalCommons@University of Nebraska - Lincoln

Library Philosophy and Practice (e-journal) | Libraries at University of Nebraska-Lincoln

Winter 2-26-2014

# Metadata for a Web Archive: PREMIS and XMP as Tools for the Task

Laurentia Romaniuk
laurentia.romaniuk@gmail.com

Follow this and additional works at: http://digitalcommons.unl.edu/libphilprac

 Part of the Archival Science Commons

Romaniuk, Laurentia, "Metadata for a Web Archive: PREMIS and XMP as Tools for the Task" (2014). *Library Philosophy and Practice (e-journal)*. 1098.
http://digitalcommons.unl.edu/libphilprac/1098

# Metadata for a Web Archive: PREMIS and XMP as Tools for the Task

Laurentia M. Romaniuk

San Jose State University

Abstract

In a time where websites are ever changing, what metadata standards and tools are best for ensuring that web archive objects (such as snapshots of websites) are readable for users of the future? Can the evolution of web interfaces be documented? Initiatives that explore these questions already exist such as the Internet Archive's Wayback Machine (which stores source code from websites along with images); however, other archive building solutions are also available but have yet to be explored. The field of digital asset management (DAM), for example, has long examined how assets (digital files) are stored, organized, retrieved, and preserved. Best practices related to the use of metadata standards and tools found in digital asset management are useful and relevant to web archive building. In order to better understand the practicality of implementing DAM best practices in building a web archive, a small project was performed which involved cross-walking two metadata standards, Adobe's eXtensible Metadata Platform (XMP) and PREservation Metadata: Implementation Strategies (PREMIS), and recording metadata related to snapshots of a website, the Perseus Digital Library, over a span of over a decade. The findings of this project showed that it is impossible, at least in part, to encode PREMIS within XMP.

*Keywords*: XMP, PREMIS, Crosswalking, Archive, Web, Interface, Digital Asset Management

# Metadata for a Web Archive: PREMIS and XMP as Tools for the Task

Websites in all their varying forms are the cheap, ephemeral, and widely used medium for communication in our age. Today much of the human cultural record is born digitally and much of the analog cultural record is being made digital through various digitization projects. Our cultural makeup is in a transition towards 0s and 1s, and these bits and bytes are largely accessible only through the web (Wijngaarden, 2007). To complicate matters, changes to websites are instantaneous, and countless changes are happening each second without anyone giving much notice. This comes as no surprise. After all, change is not simply something that people have become used to; it is something that is expected (Wijngaarden, 2007). Websites that seek to thrive online must be adaptable to user needs and feedback and so change is a necessity (Wijngaarden, 2007). Technology is transient, and those precious 0s and 1s that tell part of humanity's story need to be archived and preserved.

In a time where websites are ever changing, what metadata standards and tools are best for ensuring that web archive objects (such as snapshots of websites) are readable for users of the future? Can the evolution of web interfaces be documented? Some initiatives that explore these questions already exist such as the Internet Archive's Wayback Machine (which stores source code from websites), however, other archive building solutions are also available but have yet to be explored. The field of digital asset management (DAM), for example, has long examined how assets (digital files) are stored, organized, retrieved, and preserved. Best practices related to the use of metadata standards and tools found in digital asset management are useful and relevant to web archive building. In order to better understand the practicality of implementing DAM best practices in building a web archive, a small project which involved cross-walking two metadata

standards and recording metadata related to snapshots of a website, the Perseus Digital Library, over a span of over a decade, was performed.

## Literature Review

### Preservation Best Practices in DAM

Wijngaarden (2007) points out that while books can be left on a shelf for many years and remain readable to humans, digital objects are not human readable; instead they require the use of a machine to render digital objects legible to humans, and thus the technology that translates these objects into something comprehensible to humans must be maintained and accommodated. Because computers must read digital assets before humans can understand them, it is crucial that these assets be stored in such a way that interoperability amongst various computer types and software programs is given priority (National Science Foundaton [NSF] & Library of Congress [LOC], 2003). Cordeiro (2004) argues that in order to ensure that digital assets remain readable across platforms and thus are in alignment with preservation best practices, standardized metadata and file formats must be used, and beyond this, the metadata standards and file formats must not be proprietary; open access is crucial. Metadata standards that are used across communities and domains are more likely to remain interoperable across time, and thus selecting well established standards is ideal (Cordeiro, 2004). A key part of digital preservation best practices is to use widely accessible and well-known metadata schemas and tools when cataloguing digital objects.

### Metadata Standards for Preservation

In reviewing metadata standards for this application to a web archive project, it is important to examine standards as they pertain to preservation and also interoperability. Recommendations for digital preservation issued by OCLC (2002) state that preservation

metadata "is the information necessary to maintain the viability, renderability, and understandability of digital resources over the long term". In order to meet this definition of metadata while building a web archive, two metadata standards must be brought together; XMP and PREMIS.

*XMP*

Many members of the DAM community use Adobe's eXtensible Metadata Platform (XMP) as their primary method of cataloguing metadata (Regli, 2009). This is because XMP allows metadata to be stamped directly to assets, as opposed to having metadata assigned to assets but being stored separately from the asset (Bright, 2006). With many metadata schemas, metadata is merely associated with or refers to digital objects, and thus it risks being lost or disassociated from an object (Bright, 2006). XMP is also attractive for a DAM web archive project because it is an extension of Resource Description Framework (RDF); a standard that is used in indexing the web and digital objects (Bright, 2006; Rosziekicz, 2005). Additionally, because XMP draws from RDF which in turn draws from Dublin Core, XMP covers the majority of Dublin Core's element requirements and is relatively interoperable with the basic elements of many standards. XMP is also interoperable with eXtensible Markup Language (XML) coding, and it is open source (Binder, 2006). XMP is highly interoperable and well known, and it ensures that metadata and assets remain together.

Another strength that XMP presents in creating an archive of an ever-changing and ever-growing web is that this standard has the capacity to continually grow and adapt to a demand for increasing capacity. Bright (2006) compares XMP's structure to the web, because much like how websites found on the web link to one another and pull data from various servers and sources, so too does XMP allows users of a DAM system move asset locations and change assets formatting

and metadata freely while maintaining links between assets. XMP is a standard that mimics the web in an intrinsic sense, and because of this, it holds great potential to be used in creating a web archive.

*PREMIS*

While XMP has attributes that may be useful in creating an archive of the web, one of its limitations is that it was not designed for digital preservation initiatives. Preservation Metadata: Implementation Strategies (or PREMIS for short) is a standard that looks explicitly at ensuring that appropriate metadata is recorded to ensure that digital objects remain renderable across prolonged periods of time, and it has been widely adopted by preservationists from all areas (Donaldson and Yakel, 2012; Vermaaten, 2010). PREMIS focuses on five core elements, or entities, and they are: Intellectual (the document's representation), Object (a bitstream), Event (any changes made or events involving the asset), Agent (software, person, or other affiliated thing with asset), and Rights entities (PREMIS Editorial Committee, 2012). Each of these entities is interrelated (Figure 1). By maintaining a bitstream, a list of agents, and by knowing events that occur to an asset, proper preservation of the asset may be ensured. Fortunately, XMP – though not based on PREMIS or explicitly anchored in preservation practices – appears to accommodates all five of PREMIS' core entities (in thanks, largely, to PREMIS and XMP both having been informed in some way by Dublin Core's structure).

A possible route for developing a metadata standard to use in building a web archive then may be to use XMP with the guidance of the PREMIS data dictionary to ensure that metadata in XMP is recorded in a manner that best suits preservation needs. Kennedy (2006) notes that through understanding how each standard operates, various standards may be brought together through mixing and matching. This is often the best method for implementing metadata

standards in preservation environments (Kennedy, 2006). Thus, through mixing and matching XMP and PREMIS, a new practice for web archive building emerges.

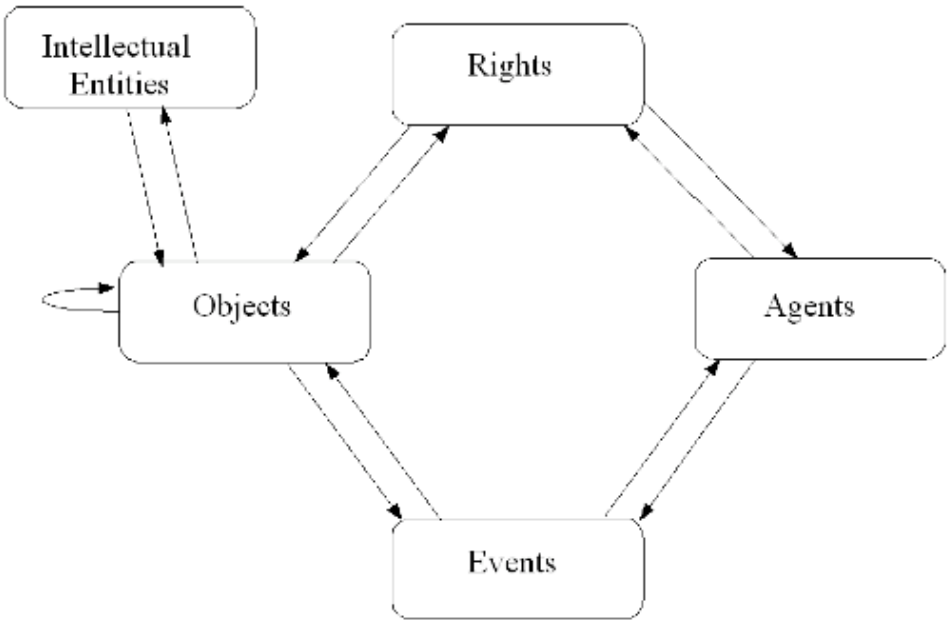


Figure 1: The PREMIS Metadata Model (PREMIS Editorial Committee, 2012).

**The Tool for the Task**

With a preliminary grasp of what metadata standards will form the base of an internet archive project, a tool to implement these standards must be chosen. Adobe has issued a software tool that is commonly used in DAM communities known as Adobe Bridge, and it allows cataloguers to easily access and change XMP metadata that is embedded into assets. Roszkiewicz (2005) explains that Bridge's interface (Figure 2) is easy enough to use that those new to cataloguing can quickly learn how assign high-quality metadata to assets. Roszkiewicz (2005) also points out that Bridge is accessible through the web and assets may be loaded into Bridge through the web. Adobe Bridge is a tool that may prove quite promising in cataloguing web images for an archive, however, in order to maintain the integrity of metadata as it is

recorded in Bridge (or in any metadata recording environment) it is crucial that cataloguers have defined rules for metadata entry as well as a controlled vocabulary.

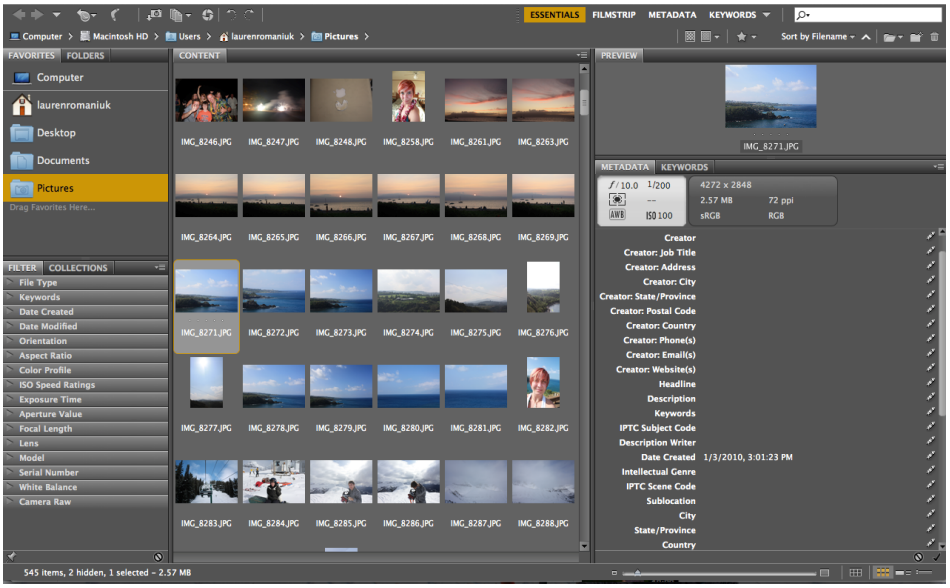


Figure 2: A screenshot of Adobe Bridge. Metadata fields are on the right and left.

**Building A Vocabulary**

Hedden (2010) explains that controlled vocabularies assist in the implementation of descriptive metadata fields as they support "consistent, accurate, and quick indexing and retrieval of digital content". Regardless of which metadata standard or combination of standards is chosen for a web archiving project, a controlled vocabulary should be provided to cataloguers in order to ensure that the metadata created throughout the project is of high quality. The same can be said when using metadata entry tools such as Adobe Bridge. Hedden (2010) also advocates for the use of a term list; such a list provides useful information such as which file formats are acceptable for a project, what usage rights are appropriate, and what the retention status of archival files should be. In order to accurately assess how successful recording metadata about websites for archival preservation is, it is crucial to use controlled vocabularies to ensure that information is recorded accurately and concisely across records.

**Case Studies**

While reviewing literature about preservation, metadata standards, and the use of vocabulary design is useful in the scoping phase of a project, so too is it important to review current case studies so as to gain a practical understanding of how projects thus far have been run.

At UPS, Barnes, Bowden, Griffith, and Keathley (2012) overhauled the DAM system to improve digital preservation practices related to archival web assets and archived client accounts because they were sunsetting various existing websites and they needed to build an archive of said sites. Through their project, Barnes et al. (2012) found that building a web archive requires the use of data dictionaries (controlled vocabularies) to ensure cataloguing success and they also concluded through their project that "one universal standard is not possible for the arrangement and description of digital objects, as no one standard would provide the specific type of access and search environment needed for all audiences". Both Kennedy's (2006) and Hedden's (2010) theories come into play here in a practical sense – blending metadata standards is a good thing in digital asset management, as is understanding audience (user) needs, and the use of controlled vocabularies.

The Colorado State University Libraries (CSUL) underwent a major digitization project in 2012 which included selecting a new metadata standard to put in place along with building a DAM. Oehlerts and Liu (2013) used the TIFF file format for archival images in this project – the same format that Cordeiro (2004) recommends – and their standard of choice was PREMIS. In order to ensure that metadata could be migrated from one DAM system to another in the future, Oehlerts and Liu (2013) maintained a record of all PREMIS metadata in a TXT file separate from the DAM that they were working with. Assets in this case could be migrated, but so too would the metadata need to be carried over in a separate form. This highlights a problem in

implementing PREMIS in DAM software and the reason why PREMIS should complement XMP, as XMP allows metadata to be affixed to assets and migrated in unison.

The Museum Victoria in Australia created a DAM system for preserving digital audio-visual files and sharing said files on the internet. Broomfield (2009) explains that XMP and Adobe Bridge were used because they have a strong presence across government and corporate archive organizations. Because XMP is largely compatible across DAM systems, the Museum Victoria was able to share their repository with other systems (Broomfield, 2009). Additionally, Broomfield (2009) paid particular attention to usage-related fields in XMP to set access privileges for assets and build metadata modules for various user groups. This use of XMP allowed users to perform their own "self-service" retrieval of digital assets, and the use of Adobe Bridge significantly improved cataloguing efficiency (Broomfield, 2009). The argument that Broomfield (2009) appears to make is that because XMP is compatible across domains, and because Bridge is easy for cataloguers to use, preservation initiatives were bolstered simply through ease of use.

Each case-study in this review highlights different areas for development and consideration in selecting and implementing a metadata standard in a DAM system, but the overall lesson in these studies appears to be this: no metadata standard is best, and attention should be given to metadata recording integrity, interoperability, and the intended use of assets in executing a DAM project. Academic and professional literature shows that it is also important to create controlled vocabularies for catalogues to ensure that cataloguers enter metadata properly, and to ensure that retrieval for end-users is optimized. Affixing metadata directly to assets has its strengths – assets can move relatively freely from one system to another without the need to keep a separate file or document with metadata. A file format for storing image-based assets such as

TIFF is ideal for metadata storage because this format allows cataloguers to embed metadata directly into assets. XMP is a useful standard, but it is not tailored specifically for preservation practices and so PREMIS, which focuses on the preservation of digital objects, should be taken into consideration when using XMP. Through understanding these points the shape for this project aimed at cataloging and archiving the web became clear.

## Description of Project

In an effort to understand how effective the application of PREMIS' data dictionary to XMP through Adobe Bridge is in building a web archive, records of various pages on a website were recorded in Adobe Bridge using PREMIS' guidelines as the backbone for cataloging. Information from the literature review leading up to this project informed the decisions made in designing this project. Screenshots from the Perseus Digital Library (Perseus) were imbued with metadata through Adobe Bridge through the use of a controlled vocabulary and cross-walking scheme to test how effective combining XMP and PREMIS together for web preservation is.

### Goals

The goal of this project was to better understand opportunities and challenges combining XMP and PREMIS together as a means for performing archival work, specifically in instances where archives of web interfaces are being stored.

Perseus (http://www.perseus.tufts.edu/hopper/) was chosen as the website of choice in this project as it is an expansive open-access tool maintained by Tufts University and it is a website that has gone through several iterations. It was a perfect fit for a project aimed at archiving a website and its various interface changes. Initially, Perseus was offered as an academic software tool available for Macintosh users. It has since gone through many web iterations, the most recent version being dubbed "Hopper". While it was only possible to take

original screen captures of Perseus as it existed at the time of this project, the Wayback Machines had replica pages of Perseus' older web interfaces available, and thus snapshots of these replica pages were used to represent archival iterations of Perseus' site. This proved useful in tracking how a website changes over time. Perseus was also an ideal candidate because it offers several different tools to its users: it has an educational section for children, research-intensive reading sections, and images and snapshots of historical places, items, or figures. An important part of archiving a website involves tracking the functionality and features of a website, and to this effect, Perseus offered a rich source of information to track in relation to usability.

**Steps & Procedure**

As discussed in the literature review, an important part of cataloguing involves providing indexers with a controlled vocabulary to work with in order to keep things consistent, as such, the first step in this project was developing a controlled vocabulary. PREMIS' (2012) data dictionary provided many very specific and useful explanations of how fields should be used and how information should be recorded, and this informed the development of a small controlled vocabulary and cross-walk for applying PREMIS to XMP. Because PREMIS does not specifically aim to record web content, fields related to web content were also added to the dictionary, along with information about certain XMP fields. The final controlled vocabulary used in this project can be found under Appendix A.

With a controlled vocabulary built, the next step of this project involved taking screenshots of Perseus to be included in a sample web archive. Twenty-five screenshots were taken in total and saved in a TIFF file format. Screenshots were selected from various areas of Perseus' site in order to capture different tools and content available to users as well as different

versions of Perseus' site in order to capture major interface changes over time. As mentioned, the only live screenshots that could be taken of Perseus were shots taken of Perseus as it existed at the time of the project. All older images are in fact screenshots of replicas of older versions of Perseus as recorded by the Wayback Machine. These Wayback Machine screen shots do not capture the screens of Perseus exactly as they were at the time they were taken, but they are the best available examples of what Perseus looked like at that time available. Ideally, screenshots of a website should be taken of a website as it exists in its present state over a period of years to build a web archive, but this was not possible given the time constraints of this project.

Once screenshots were collected, they were imported into Adobe Bridge, and metadata was assigned to them as per the controlled vocabulary guidelines. While entering PREMIS metadata in XMP through Bridge, notes were recorded by the cataloguer (the author of this paper) regarding any difficulties or concerns that arose in the cataloging process. Information from these notes was then used to determine the effectiveness of combining these two schemas to record preservation metadata for web-based assets.

## Product and Results

Twenty-five TIFF (.tif) files and twenty-five XMP (.xmp) files make up the product of this project. Once TIFF files were imbued with metadata through the cataloging process, their metadata was exported through Adobe Bridge to create the XMP files. A controlled vocabulary and cross-walk for implementing mandatory PREMIS fields in XMP through Adobe Bridge was also created.

In addition to the controlled vocabulary and cross-walk for this project, specifications about how to take snapshots of Perseus should have been outlined. There are various tools available to take screenshots of websites, and some are better than others. Better models tend to

take a picture of the whole page, including elements that are not immediately visible in the browser, while some only take snapshots of what is visible above the fold. The screenshot software used in this project was of a poor quality, and so the size and content of the images is not as consistent as it may have been with clear guidelines and a high-quality software tool. Another guidelines-related challenge was that the URL of websites sometimes changed entirely across iterations – while the content of one page clearly was an evolution of an older one, the site's overall architecture changes often meant that the URL between these two pages were not at all similar. How does this effect determine which page is the iteration of another? Guidelines on determining what counts as an "iteration" are needed. Ideally, a large web archiving project would have clear best practices for collecting asset information. As this project was more concerned with implementing metadata rather than asset collection, it was not within the scope of this project to create detailed acquisition and preservation guidelines.

With regards to creating the cross-walk specifically, it was relatively easy to match PREMIS fields to XMP, indeed because of the reason mentioned in the literature review of this report – both PREMIS and XMP have a history that traces back to DublinCore. That being said, some issues with regards to cross-walking did arise. PREMIS' data dictionary has been written under the assumption that metadata is not affixed directly to assets in the way that XMP accommodates. As a result of this, PREMIS has many required fields that more or less describe the relationship between certain metadata fields and other fields or the asset itself. These fields are not necessary when the asset holds the metadata as it did in this project, as the metadata's relationship to the asset is implicit. One required field that was not listed as part of the metadata in Adobe Bridge was what PREMIS refers to as inhibitorType. This field handles encryption elements for the file. Adobe Bridge manages this information separately – it is not metadata that

a cataloguer records in the metadata fields, but rather, it is an option that may be toggled on or off. Last, the one overall concern in cross-walking PREMIS into XMP through Adobe Bridge is that XMP caters primarily to representation and file metadata, while PREMIS also asks that metadata be recorded in relationship to the bitstream of the asset. This metadata is automatically stored within a .tif file, but it is not editable through XMP in Adobe Bridge. Thus, only two out of three areas of metadata management outline in PREMIS were touched upon.

With regards to the controlled vocabulary for this project, the naming convention was a bit redundant too. The first element of the naming convention consisted of coding the date that the snapshot was taken from. However, this metadata is automatically recorded when a snapshot is taken and a file is created. When snapshots of a site are taken in of the site as it exists in present day, this element is not needed. For this project, because snapshots were being taken of archival mirrors of Perseus at different points in time, then the snapshot that was taken was not synonymous with the date of the website. In a web archive project where images are being taken on a regular basis of present-time web pages, this date-element of the naming convention is not needed.

One final take-away from this project had to do specifically with the usability of any given web page. If one is applying PREMIS metadata alone to XMP, then the metadata that is collected with be purely preservation metadata. This metadata is useful for preserving assets, however, it is not useful in describing how an asset functions or what its evaluative value is in contrast to other assets in a collection. Adobe's XMP accommodates a wide variety of metadata, and so metadata related to the usability of a webpage could perhaps be recorded in those fields. That being said, one useful bit of metadata to record for a snapshot of a website would be the site's code. The Wayback Machine captures this metadata, while a snapshot alone does not. This

code should also have been captured and affixed to the asset in some way through XMP to ensure that the structure and usability of the website was articulated in some way through the metadata.

Overall, this project was a success, despite the observations noted herein. PREMIS metadata absolutely can be implemented in XMP, particularly through Adobe Bridge, so long as cataloguers have a cross-walk schema available. Future projects can perhaps be informed by observations coming out of this report to further explore the use of PREMIS and XMP in web archiving initiatives. Certainly, there is potential for a PREMIS and XMP based web archive to work well, but the project must consider all elements of preservation, not simply metadata, and the project must occur over a very long time period.

## Conclusion

Although PREMIS and XMP may be combined effectively, there remains no clear-cut metadata standard or method for building a web archive, and nor should there be. Oehlerts and Liu (2013), Barnes et al., (2012), and Broomfield (2009) all found in their case-studies that a clearly needed part of any preservation project or asset management project involves continual evaluation and auditing to understand the effectiveness of any metadata initiative. This metadata project provided critical observations for a new metadata initiative with promise to take advantage of many well-established tools in the field of DAM, but much like the ephemeral content which a project of this nature tries to capture, so too must the initiative be flexible enough to adapt and change.

## Afterword

I would like to express thanks to Dr. Geoffrey Rockwell and the Implementing New Knowledge Environments (INKE) team at the University of Alberta for introducing me to the

Perseus Digital Library and also, for teaching me to think critically about interfaces for academic research, as it is through working with this team that my interest in the topic of web archiving began. Also, thank you to Dr. Mary Bolin for prompting me to do this research project through her LIBR 281 – Metadata course at San Jose State University, and her encouragement of me submitting this research for publication.

# References

Barnes, M. K., Bowden, F., Griffith, J., & Keathley, E. F. (2012). The sorting of competing metadata models in digital asset management. *Journal of Digital Media Management*, *1*(1), 78-84.

Binder, J. (2006). Exchanging assets and metadata across platforms. *Journal of Digital Asset Management, 2*(5), 215-218.

Bright, J. (2006). First steps: XMP. *Journal of Digital Asset Management, 2*(3/4), 198-202.

Broomfield, J. (2009). Digital asset management case study: Museum Victoria. *Journal of Digital Asset Management, 5*(3), 116-125.

Cordeiro, M. I. (2004). From rescue to long-term maintenance: Preservation as a core function in the management of digital assets. *VINE, 34*(1), 6-16.

Donaldson, D. R., & Yakel, E. (2012). Secondary adoption of technology standards: The case of PREMIS. *Journal of Archeological Science, 13*(55), 55-83. DOI: 10.1007/s10502-012-9179-0

Kennedy, D. (2006). Metadata standards: The DAM accelerator! *Journal of Digital Asset Management, 2*(3/4), 134-137.

Hedden, H. (2010). Taxonomies and controlled vocabularies best practices for metadata. *Journal of Digital Asset Management, 6*(5), 279-284.

NSF & LOC. (2003). It's about time: Research challenges in digital archiving and long-term preservation, final report. Retrieved from http://www.digitalpreservation.gov/documents/about_time2003.pdf

OCLC. (2002). Preservation metadata and the OAIS information model: A metadata framework

to support the preservation of digital objects. Retrieved from http://www.oclc.org/content/dam/research/activities/pmwg/pm_framework.pdf?urlm=161391

Oehlerts, B., & Liu, S. (2013). Digital preservation strategies at Colorado State University libraries. *Library Management, 34*(1/2), 83-95.

PREMIS Editorial Committee. (2012). *PREMIS data dictionary for preservation metadata, version 2.2*. Washington, DC: Library of Congress. Retrieved from: http://www.loc.gov/standards/premis/

Regli, T. (2009). The state of digital asset management: An executive summary of CMS watch's digital asset management report. *Journal of Digital Asset Management, 5*(1), 21-26.

Roszkiewicz, R. (2005). Upstream metadata and the impact of the IPTC schema for XMP and Adobe Bridge. *Journal of Digital Asset Management, 1*(4), 298-302.

Vermaaten, S. (2010). A checklist and a case for documenting PREMIS-METS decisions in a METS profile. *D-Lib Magazine, 16*(9/10).

Wijngaarden, H. (2007). Long-term preservation and permanent access: How to ensure the long-term reuse value of your digital assets. *Journal of Digital Asset Management, 3*(2), 102-109.

Appendix A
Controlled Vocabulary and Cross-Walking Guide

**File Format**

All files should be stored as TIFF (.tif) files.

**File Naming Convention**

All files names must follow this naming format: yymmdd_sitename_pagename.tif. In this instance, the year, month, and date represent the date that the screenshot of the web page was taken. The sitename should be the prescribed sitename listed in this controlled vocabulary (see below). The pagename should be a descriptive title that is derived from the title of the webpage in the screenshot. For example, a webpage with the title "Our Research" could be recorded as "Research". The term "Research" would then be used across iterations of a page.

*Sitenames* – A sitename should be chosen for each website recorded. In this project, only one website was accessed (the Perseus Digital Library). The sitename assigned to this website was "Perseus".

**Cross-Walking PREMIS and XMP in Adobe Bridge**

The following tables present cross-walking paths for implementing the metadata required by the *PREMIS Data Dictionary for Preservation Metadata Version 2.2* (PREMIS Editorial Committee, 2012) into XMP as it is stored in Adobe Bridge. For specific information about what each field represents or how metadata should be recorded, review the *PREMIS Data Dictionary for Preservation Metadata Version 2.2*

In instances where the XMP value is listed in quotations (""), the field name in and of itself is the exact value in XMP that meets the parameters of PREMIS' required field. When an XMP value is not listed in brackets the field's metadata must be entered by the cataloguer as per the guidelines found in the *PREMIS Data Dictionary for Preservation Metadata Version 2.2*

| PREMIS | XMP | Rules for Project |
|---|---|---|
| **agentIdentifierType** | "Creator"; "Description Writer" * | |
| **agentIdentifierValue** | Creator; Creator: Job Title; Creator: Address; Creator: City; Creator: State/Province; Creator: Postal Code; Creator: Country; Creator: Phone(s); Creator: Email(s); Creator: Website(s) | |
| **compositionLevel** | Instructions | Enter: "Composition Level: none" |

| | | |
|---|---|---|
| **copyrightStatus** | Copyright Status | |
| **copyrightJurisdiction** | Copyright Notice | Enter: “Jurisdiction: U.S.A.” |
| **dateCreatedByApplication** | Date Created | |
| **eventIdentifier** | “Date File Modified” | |
| **eventDateTime** | Date File Modified | |
| **format** | “Document Type” | |
| **formatName** | Document Type | |
| **formatRegistryName** | Instructions | |
| **formatRegistryKey** | Instructions | Enter: “TIFF Registry: Adobe - http://partners.adobe.com /public/developer/tiff/index.html” |
| **hwType** | Description | Enter: “Hardware Type: Intel Processor” |
| **objectCharacteristics** | Color Mode; Color Profile | |
| **objectIdentifier** | Filename | |
| **objectIdentifierType** | Job Identifier | Enter: “Website Snapshot (See Source)” |
| **objectIdentifierValue** | Source | |
| **rightsStatement** | Rights Usage Terms | Enter: “Rights Statement: Asset may be used under Fair Use policy.” |
| **rightsBasis** | Rights Usage Terms | Enter: “Rights Basis: Asset created in U.S.A. and thus Fair Use applies.” |
| **significantPropertiesType** | “Dimensions” | |
| **significantPropertiesValue** | Dimensions | |
| **size** | File Size | |
| **swType** | Description | Enter: “Software Type: Image Viewer” |

* In this instance, “Description Writer” refers to the name of the cataloguer recording metadata. This field should abide by the following format: Firstname Lastname