Process history metadata for
time-based media artworks at the
Museum of Modern Art, New York

Received: 31st January, 2016

Peggy Griesinger
is the Metadata and Cataloguing Librarian at George Mason University Libraries. She is responsible for the copy and original cataloguing of audiovisual materials and consults throughout the library on metadata projects. Prior to this, she was the National Digital Stewardship Resident at the Museum of Modern Art in New York City.

George Mason University Libraries, 4400 University Drive, MS 2FL, Fairfax, VA 22030, USA
Tel: +1 703 993 8948
E-mail: mgriesin@gmu.edu

Abstract The Museum of Modern Art in New York City (MoMA) is a world-renowned museum with a large collection of time-based media artworks. Media conservators in MoMA’s conservation department noticed a gap in their documentation for time-based media artworks: there was no standardised way of recording the digitisation or format migration history of an artwork, known as its process history. Conservators require such information to store and exhibit an artwork in a way that matches the artist’s original vision. The National Digital Stewardship Resident at MoMA was tasked with creating a standardised metadata profile that could record and store this information. The resulting profile uses the PREMIS, PBCore, reVTMD and METS schemas to record technical metadata about every tool used to digitise or migrate an artwork from one format to another. This metadata profile ensures that the artistic and bit-level authenticity of time-based media artworks can be maintained and understood now and in the future.

KEYWORDS: metadata, digital preservation, time-based media, art conservation

INTRODUCTION
The Museum of Modern Art in New York City (MoMA) has one of the most important collections of modern art in the world. Perhaps less famous than its collection of Picassos and Van Goghs is its extensive collection of time-based media artwork, which is defined by the Guggenheim Museum as ‘contemporary artworks that include video, film, slide, audio, or computer-based technologies’.1 MoMA has formally collected time-based media artworks since the founding of its Film Library in 1935, a collection which now houses more than 20,000 films.2 Over the past decades, MoMA has extended this collection to include a wide array of artworks that make use of the possibilities provided by time-based media.

Over the years, conservators have digitised and migrated this collection to different physical and digital formats. Media conservators in the conservation department at MoMA must ensure that these artworks, in their various formats, are preserved, both in terms of bit-level authenticity and artistic authenticity. However, media conservators noted a gap in their preservation of these...
materials: a lack of standardised metadata to describe the digitisation or migration history of these items. This paper refers to this digitisation and migration history as ‘process history’. The paper will detail the project undertaken at MoMA, in conjunction with the National Digital Stewardship Residency in New York (NDSR-NY), to research appropriate metadata standards to resolve this issue and create a metadata profile for use in MoMA’s existing digital repository for time-based media artworks.

MOHAM’S DIGITAL REPOSITORY FOR TIME-BASED MEDIA

The media conservation team at MoMA manages the museum’s collection of time-based media artworks using an open-source system called Binder. Binder builds on the open-source Archivematica and Access to Memory (AtoM) software to create a digital repository management interface specifically designed for cultural heritage institutions to manage their time-based media collections. The metadata profile described in this paper is intended to be integrated into this existing system. Binder has a number of aspects pertinent to this project, including the ability to view extensive technical metadata alongside an object, the ability to visualise the relationships between different iterations of an artwork (i.e., original tape, preservation copy, viewing copy), and the ability to search and browse based on metadata associated with artworks. However, the functionality to integrate the process history metadata into the existing framework was not present, and the desire to take advantage of these functions for the process history metadata informed the development of the metadata, as this paper will discuss in further detail.

RESEARCHING THE METADATA PROFILE

The goal of the metadata profile for this project was to describe what is referred to at MoMA as the process history of an artwork. The term ‘process history’ encompasses the wide variety of actions that can be taken upon an artwork and the complex path an artwork can travel to end up in the form or forms currently held by MoMA. Figure 1 provides an example of how one artwork could move between different formats, and the arrowed lines indicate the relationship that conservators wished to describe — in far greater detail than that represented in

Example artwork with iterations

![Diagram of an artwork with iterations]

Figure 1: An example of an artwork and the various formats to which it could be migrated. Each arrowed line represents the process history that leads from one iteration of the artwork to another. The metadata profile for this project describes the granular detail of these process histories.
Process history metadata for time-based media artworks

this figure — using process history metadata. Archivematica already records the portion of this history that occurs during ingest into the repository, such as virus checking and file-type identification. The goal of this profile, therefore, is to create a standardised profile for describing what happens to these artworks before ingest into the repository and, on some occasions, even before being acquired by the museum.

Conservators needed a way to describe every single tool that was used to transform an artwork from one format to another. Beyond just uniformly naming these tools, the conservators required a metadata profile that could record the technical details of these tools and their role in the process history. The metadata profile needed to record basic technical metadata about the tool, how exactly that tool was used in relation to a particular artwork during a digitisation or migration process and information about the agent responsible for the process. For example, conservators needed to describe not only that capture software was one of the tools used to transform the artwork in Figure 1 from a U-matic tape to a digital file, but also the model, manufacturer, serial number and settings of that capture software, as well as the name of the conservator who initiated the capture.

The first phase of building the metadata profile involved interviewing media conservators and becoming familiar with MoMA's time-based media artwork collection in order to conduct a needs assessment. This required a review of existing documentation for time-based media artworks. The findings showed that media conservators made attempts to record process history metadata, but that these metadata were not standardised and were not being stored in a format that would allow for long-term preservation. Information tended to be recorded as free text in Microsoft Word documents without the use of controlled vocabularies. Furthermore, this information could sometimes be difficult to locate within the local file-storage systems at MoMA. Conservators expressed a need for a metadata profile that would make it simpler for them to record and access process history information.

The next phase of developing the process history metadata profile required research into existing metadata standards that could be used as a solution for MoMA's needs. This began with a survey of existing research that documents the variety of standards available for cultural heritage institutions. Jenn Riley's ‘Seeing Standards: A Visualisation of the Metadata Universe’ was particularly helpful. Although this was published in 2010 and thus does not include the most recent standards, it was nonetheless extremely useful in gaining a sense of how existing standards relate to each other and content domains.

The vast majority of this research phase involved reading the documentation accompanying standards wherever available. This included examining XML schemas, best practices guides, example records and use cases. The author only considered standards that were reasonably documented and related in some way to digital preservation, audiovisual materials, museums or fine art. This section will detail a number of standards examined for this project, and the pros and cons of each.

Preservation Metadata: Implementation Strategies (PREMIS) is a widely adopted and supported standard for preservation metadata. It allows for the description of four entities: objects, events, agents and rights. PREMIS is already used extensively in MoMA's digital repository — Archivematica records technical metadata about the digital objects, as well as events that happen during ingest, such as checksums and file-type identification. Binder then leverages the PREMIS metadata for management purposes in the repository. Although PREMIS can describe events in a digital object's history, it does not have the level of granularity required for recording process
history events. Furthermore, the PREMIS metadata being generated by Archivematica could not be applied to events that happen before ingest or to physical iterations of the artworks. The metadata do, however, allow for XML extensions, which allow for the insertion of records written in other standards into a PREMIS record for granular description of process history. This benefits the metadata profile by allowing for the use of a widely adopted and supported standard like PREMIS, while still allowing for the level of detail necessary for the metadata profile to be successful.

The Public Broadcasting Metadata Dictionary (PBCore) is a standard developed to describe audiovisual collections, specifically in the context of public broadcasting organisations. It allows for the description of the intellectual content of an audiovisual object, and then allows for the technical description of ‘instantiations’ of that object. Instantiations refer to different versions of the same intellectual work. So, a user could describe both the VHS and WAV file copy of an episode of ‘Sesame Street’, within the same record. PBCore’s ability to handle both physical and digital objects made it especially pertinent for this project. The conservators at MoMA expressed a desire to be able to describe not only the process history that led to the creation of digital files, but also the process history that led to different versions of physical, mostly analogue, iterations of an artwork. The other standards examined in this project did not take physical objects into consideration; so any metadata pertaining to the physical aspect of an iteration of a work would require PBCore elements to be described properly. PBCore was, however, designed to record the intellectual and technical metadata relating to those objects; it cannot describe events that happen in the life of an artwork or a digital object. PBCore filled a gap for physical-only objects, but it did not answer the need for detailed process history metadata.

reVTMD is a standard developed by the National Archives and AVPreserve for the purpose of documenting process history relating to digital objects. It is, therefore, an obvious and important candidate for use in the metadata profile. The fact, however, that it is neither well documented nor widely adopted concerned the project team. The project team attempted to adhere as strongly as possible to digital preservation best practices. Therefore, the metadata profile needed to be made up of robust standards with proven track records. Because nothing else existed to fill this integral gap in the profile, however, reVTMD would need to be used.

Metadata Encoding and Transmission Standard (METS) is a widely supported and adopted standard that records technical, administrative and digital provenance metadata related to digital objects. More importantly, it can be used to link different metadata standards together into a semantically coherent document. In addition, Archivematica uses METS already, and Binder leverages METS to organise the existing technical metadata about digital objects.

A number of other standards for inclusion in the metadata profile were considered and ultimately rejected. These included a large number of standards designed to record information about art museum collections. Research into standards including Categories for the Description of Works of Art (CDWA) and Lightweight Information Describing Objects (LIDO) demonstrated that these standards were designed around more traditional artwork types, such as painting and sculpture. Although they had the necessary structure to record extensive information about those media, they lacked the granularity to describe time-based media artworks in any level of useful detail.

THE METADATA PROFILE
The final metadata profile accepted for this project uses METS to wrap multiple sections...
of technical and digital provenance metadata, which are recorded using PREMIS, PBCore and reVTMD. The reVTMD section contains the process history, which is the main content of the profile. Because reVTMD does not yet have a large and active user community, it was decided that reVTMD should be used only as an extension of more well-established standards, despite its prominence in the content of the metadata records. The metadata profile is presented in Appendix A, and a simplified diagram of the profile is shown in Figure 2.

The organisational structure of a reVTMD record allows for the description of a single digital object at a time. Extensive technical metadata about that object can be recorded; for the purposes of this project, however, those fields were not used because Archivematica already generates this type of technical metadata during ingest of the digital object into the repository. reVTMD is important and useful because of its ability to describe what it calls ‘capture history’.

The capture history element has sub-elements that record information such as date of digitisation and digitisation engineer. The repeatable wrapper element ‘coding process history’ (CPH) is the most important aspect of this, and it is the heart of the metadata profile devised for this project. Within every capture history element, any number of CPH wrappers can be recorded, each documenting a different device used in the process history of a digital object. Each CPH wrapper has sub-elements that can be used to describe these devices in technical detail. These sub-elements include role, manufacturer, model name, serial number, signal, settings and a number of other elements. This makes it possible to record not only that a playback device was used to digitise an artwork, but also granular technical detail about that device. reVTMD is the only existing standard that allows for such a level of detail when describing process history.

The media conservators expressed a desire to include process history as part of the

---

**Figure 2:** The basic sections of the metadata profile; the arrowed lines indicate nested relationships.
chain of events already attached to digital objects by the METS files generated by Binder upon ingest into the repository. This influenced the decision to use PREMIS as part of the metadata profile. If the process history could be placed within a PREMIS event element, then it would make it much easier to design functionality that would allow the process history to be viewed as another step alongside other PREMIS events such as ingest, file validation and virus checking. Media conservators would be able to view the entire recorded history of an artwork, from its creation to its migration into MoMA-approved formats; all of this information could be parsed from PREMIS event metadata, rather than needing to be gathered from disparate — and sometimes conflicting — sources. PREMIS also offered longevity and support as a standard — it has a highly active community, and the PREMIS Editorial Committee was writing the third version of the PREMIS data dictionary (PREMIS 3) as this metadata profile was being created; so it was clearly still being well maintained.

The process history metadata profile uses the PREMIS event, agent and object intellectual entities. PREMIS object records the basic identifiers for the artwork, such as file name or MoMA accession number. PREMIS event describes the process history as a distinct event with its own unique identifier. PREMIS event records the date of a process history event as well as linking it to the PREMIS agent involved with that event. In this case, an extension allows the environment element into the agent wrapper; so the agent in this case is actually the environment of the process history. Environment can include an extension; so this is where reVTMD is inserted into the PREMIS record. The reason for this seemingly convoluted method of extending PREMIS is because the metadata record makes semantic sense, as the agent responsible for the process history is actually the environment that enabled it to happen. For example, when a 16mm film is digitised, the agent responsible for this digitisation event is the entire process history environment of tools used to make the digitisation possible. In this case, PREMIS has no other way of recording an environment as an agent; so in order to record the detailed metadata necessary for describing a complete process history, the metadata profile requires multiple extensions. The recently released PREMIS 3 may eliminate the need for multiple extensions, but it was not released before the completion of the metadata profile.

PBCore primarily adds any other information that is not available in the other standards, especially concerning formats that are physical only and cannot be run through programs such as MediaInfo that automatically extract technical metadata from digital files. PBCore is the only one of the surveyed standards that, for example, allows for the recording of the physical type of the item. PBCore, therefore, became very important when conservators wanted to describe a migration from film to U-matic tape, for example.

METS is the standard that ties all of these other standards together. The PREMIS object and PBCore sections are all considered technical metadata (techMD) sections of a METS record. The PREMIS event and PREMIS agent sections are considered digital provenance metadata (digiprovMD) sections of a METS record. This allows for each section not only to be easily identified and accessed within the record, but it also more clearly delineates the role of each of these standards in the entire metadata profile.

To summarise this profile, the document itself is a METS Extensible Markup Language (XML) record. The METS record has four parts: PREMIS object, PBCore, PREMIS event and PREMIS agent. Within PREMIS agent, the XML extension includes a reVTMD object that records the process history itself.
HOW METADATA WILL BE IMPLEMENTED IN THE REPOSITORY

With a metadata profile as complex as the one created for this project, the media conservators needed to be presented with a much simpler system for entering the data to create the actual records. This required the development of extensive use cases to determine how exactly the conservators foresaw using the metadata in Binder. Interviews with the conservators revealed the necessity of, first and foremost, the ability to create and view process history metadata. However, conservators also expressed a desire to be able to perform faceted browses using criteria from the process history metadata, to compare the process history metadata of various iterations of an artwork side by side, and to perform advanced searches using elements from the process history metadata alongside the pre-existing technical and descriptive metadata.

Use cases were developed to reflect these needs, and it was determined that the conservators would create process history records using web forms with free text as well as controlled term lists. The conservators could view these records in a web browser without downloading an XML file, using the same Binder functionality that displays select technical metadata when viewing a digital object in the repository; that is, the selection of pertinent elements that are then translated into human-friendly text and displayed in a sidebar associated with an artwork. Similarly, developers could potentially use Binder’s existing functionality to allow users to view multiple iterations of an artwork, side by side, for comparison purposes, and also design advanced search functionality. Documentation was drawn up for the use cases and potential solutions; however, that functionality has not yet been created and integrated into Binder.

CONTROLLED VOCABULARIES

Controlled vocabularies were created for this metadata profile so that proper browsing and searching functionality could be ensured. There are already a number of existing controlled vocabularies that relate to audiovisual objects and artworks. For example, both PBCore and the European Broadcasting Union Core Metadata Set (EBUCore) have extensive controlled vocabularies relating to everything from video compression types to the appropriate terms to use to designate if something is colour or black and white. Searches, however, uncovered no existing vocabularies to describe the model and manufacturer name of audiovisual preservation tools. Such a vocabulary needed to be generated in conjunction with the conservators to accurately reflect the tools they were using currently in the lab, and to instruct the conservators on how to add new tools to the vocabulary as they would be introduced to the lab in the future. This was completed by conducting interviews with the conservators, as well as an examination of a list of items purchased for the media conservation lab provided by MoMA. A controlled vocabulary of current media conservation lab devices was created alongside guidelines for adding items to the list as necessary. These guidelines were part of a larger best practices document that guides the conservators in the use of all the elements editable by them in the metadata profile.

FURTHER WORK AND LESSONS LEARNED

The importance of collaboration with domain experts became clear when attempting to create a metadata profile during this project. Although the author had the expertise to design a metadata profile and evaluate existing standards, none of that work would have been of nearly as high a quality, nor as useful to the cultural heritage community, without the invaluable input of the media conservators at MoMA. Their input guided this metadata profile to a place that would genuinely serve the...
needs of the conservation community. It also ensured that there were domain experts on hand to correct any decisions in the metadata profile that would conflict with existing conservation practices or simply not be reflective of the needs that inspired this project.

The project also shone a light on the intense complexity of developing software functionality. The design of the use cases documentation took much longer than anticipated, and the level of complexity in the proposed solutions to these use cases was much higher than what had been anticipated at the outset of the project. One should never take for granted how much work can go into making what appears on the surface to be a small or insignificant change to software functionality.

One large gap that became apparent during the course of this project is the lack of metadata standards for describing the physical layout and construction of exhibits of specific artworks in museums and other spaces. During interviews and informal discussions with conservators, they expressed an interest in a method for recording this type of information in a standardised format that was easily accessible and storable. This is especially relevant to time-based media artwork, as the installation of audiovisual media can greatly complicate an already highly complex process. Conservators are already recording this information in some cases, but it lacks standardisation and accessibility. This would be a very useful area to explore for the development of future standards.

CONCLUSION

The metadata profile described in this paper offers a method for interested parties to record the process history of their audiovisual materials using metadata standards. Although it was designed with time-based media art in mind, it can be extended or simplified to describe the digitisation or format migration of almost any time-based media assets. The importance of recording this information cannot be overstated; without knowing how time-based media assets were transformed and potentially altered throughout their history, the forensic and artistic authenticity of these assets cannot be assured. This metadata profile offers a solution for ensuring that the history and provenance of these pieces can be described, stored and accessed today and into the future.

References

Appendix A: Metadata profile

```xml
<?xml version='1.0' encoding='UTF-8'?>
<mets xmlns='http://www.loc.gov/METS/' xmlns:xsi='http://www.w3.org/2001/XMLSchema-instance'
xmlns:xlink='http://www.w3.org/1999/xlink' xmlns:dcterms='http://purl.org/dc/terms/'
xmlns:premis='info:lc/xmlns/premis-v2'
xmlns:pbcore='http://www.pbcore.org/PBCore/PBCoreNamespace.html'
xlns:revtmd='http://nwtssite.nwts.nara/schema/
<amdSec>
<techMD ID='techMD_001'>
<mdWrap MDTYPE='PREMIS:OBJECT'>
<xmlData>
<premis:object xsi:type='premis:file'>
<premis:objectIdentifier>
<premis:objectIdentifierType/>
<premis:objectIdentifierValue/>
</premis:objectIdentifier>
<premis:objectCharacteristics>
<premis:compositionLevel/>
<premis:format>
<premis:formatDesignation>
<premis:formatName/>
</premis:formatDesignation>
</premis:format>
<premis:formatRegistry>
<premis:formatRegistryName/>
<premis:formatRegistryKey/>
</premis:formatRegistry>
</premis:format>
</premis:objectCharacteristics>
<premis:linkingEventIdentifier>
<premis:linkingEventIdentifierType/>
<premis:linkingEventIdentifierValue/>
</premis:linkingEventIdentifier>
</xmlData>
</mdWrap>
</techMD>
</amdSec>
</mets>
```
Process history metadata for time-based media artworks

<premis:agentIdentifierType></premis:agentIdentifierType>
<premis:agentIdentifierValue></premis:agentIdentifierValue>
</premis:agentIdentifier>
<premis:agentType></premis:agentType>
<premis:agentExtension>
<premis:environment>
<premis:environmentPurpose></premis:environmentPurpose>
<premis:environmentExtension>
<revtmd:reVTMD>
<revtmd:object>
<revtmd:filename></revtmd:filename>
<revtmd:identifier></revtmd:identifier>
<revtmd:mimetype></revtmd:mimetype>
<revtmd:captureHistory>
<revtmd:digitizationEngineer></revtmd:digitizationEngineer>
<revtmd:role></revtmd:role>
<revtmd:description></revtmd:description>
<revtmd:manufacturer></revtmd:manufacturer>
<revtmd:modelName></revtmd:modelName>
<revtmd:serialNumber></revtmd:serialNumber>
<revtmd:signal></revtmd:signal>
<revtmd:settings></revtmd:settings>
</revtmd:codingProcessHistory>
<revtmd:codingProcessHistory>
<revtmd:role></revtmd:role>
<revtmd:description></revtmd:description>
<revtmd:manufacturer></revtmd:manufacturer>
<revtmd:modelName></revtmd:modelName>
<revtmd:serialNumber></revtmd:serialNumber>
<revtmd:signal></revtmd:signal>
<revtmd:settings></revtmd:settings>
</revtmd:codingProcessHistory>
<revtmd:codingProcessHistory>
<revtmd:role></revtmd:role>
<revtmd:description></revtmd:description>
<revtmd:manufacturer></revtmd:manufacturer>
<revtmd:modelName></revtmd:modelName>
<revtmd:serialNumber></revtmd:serialNumber>
<revtmd:signal></revtmd:signal>
<revtmd:settings></revtmd:settings>
</revtmd:codingProcessHistory>
<revtmd:codingProcessHistory>
<revtmd:role></revtmd:role>
<revtmd:description></revtmd:description>
<revtmd:manufacturer></revtmd:manufacturer>
<revtmd:modelName></revtmd:modelName>
<revtmd:serialNumber></revtmd:serialNumber>
<revtmd:signal></revtmd:signal>
<revtmd:settings></revtmd:settings>
</revtmd:codingProcessHistory>
</revtmd:captureHistory>