Presented at the Electronic Media Group Session, AIC 35th Annual Meeting May 15–20, 2009, Los Angeles, CA.

A CASE FOR DIGITAL CONSERVATION REPOSITORIES

BARBRA MACK AND GLENN WHARTON

ABSTRACT

This article proposes a broad model for conservation documentation and information management of digital artworks in museums. Through analysis of a work at the Museum of Modern Art, New York City, the authors argue for a multi-level approach that honors the values assigned by the artists along with the requirements of underlying technologies. Justification is presented for a managed conservation repository that maintains digital assets and associated technical metadata, together with information about relationships between components as they change over time. The repository model is a non-hierarchical, component-based information management system.

INTRODUCTION

The conservation and management of complex digital artworks requires a manifest approach, with a focus on maintaining the artists' vision while developing different technical strategies for each component. The term *manifest* places attention on maintaining the intended viewer experience, even as some of the technology changes in order to keep the work operative. The delicate balance of managing change while respecting artist-defined integrity is compounded when individual components require different value assessments and conservation strategies.

This case study research was undertaken in the context of a larger collaboration between the Museum of Modern Art (MoMA) and New York University (NYU). MoMA has over four hundred works with digital components in its collection. While most

of the works are digital video, approximately twenty are complex computer-driven installations. In 2009, NYU created a Conservation of Computer-Based Art (CCBA) working group composed of faculty, staff, and graduate students from departments across the university.[1] Through this group, the university and the museum are developing new research, curricula, and educational opportunities, while modeling new programs for conserving digital art.

The artwork, *I Want You to Want Me (IWYTWM)* (2008), by Jonathan Harris (b. 1979) and Sep Kamvar (b. 1977), provides a particularly rich case to frame the conservation needs of complex digital art. It employs a multi-component networked system, which is an increasingly common practice among contemporary media artists. Exploring the related complexities of this practice provides a tangible case to understand why its conservation requires a variable approach based on artist concerns and technical requirements. It also forces recognition that conservation must include a managed digital conservation repository, which employs a multi-component object model to document technical metadata and insure integrity of the digital assets.

IWYTWM is a born-digital, interactive, installation that explores the phenomena of online dating (Harris and Kamvar 2008). MoMA curator, Paola Antonelli, of the Architecture and Design Department commissioned it in 2008 for the *Design and the Elastic Mind* exhibition from the artistic partners Jonathan Harris and Sep Kamvar (MoMA 2008). The public interacts with a touch-screen that generates five movements: "Who I Am," "What I Want," "Snippets," "Matchmaker," and "Breakdowns." Within each movement, the viewer refines searches for data that are harvested from dating websites. For instance, touching pink and blue balloons on the screen produces still images, videos, and text from online posts about who people are and what they are looking for in a date.

The authors interviewed artists Jonathan Harris (Harris 2008) and Sep Kamvar (Kamvar 2008) as part of this research. Deena Engel, Department of Computer Science, NYU, and Mona Jimenez, Moving Image Archiving and Preservation, NYU, as part of a CCBA meeting at NYU, interviewed Sep Kamvar again (Kamvar 2010). In addition, Deena Engel performed a risk assessment of the work that details conservation risks associated with the computer technologies employed by the artists (Engel 2010).[2]

A WEB OF COMPONENTS: THE ABSTRACT IT SYSTEM

IWYTWM is comprised of a number of hardware and software components using varying technologies. The work was networked during the first exhibition to query data that was harvested live from over twenty dating web sites. A state-of-the-art, high-resolution, custom touch-screen panel was selected to display slowly descending balloons that contain text statements from individual profiles taken from the sites. A person viewing the work may interact with the touch screen by selecting from one of the five movements (figs. 1, 2).

As a complex, multi-component digital artwork, *IWYTWM* provides an apt case to explore the need for a managed digital repository. Some of the components were custom made by the artists, some were custom manufactured for the artists, and some were commercially obtained. Due to the variation, each component requires a different level of conservation and dictates varying conservation practices. The primary components that comprise the work include the following:

- C++ custom application
- Touchscreen display, high-resolution (Tek Panel)
- Java web crawler
- IWYTWM dataset
- MySQL database application
- Server hardware on which the Java Web crawler and MySQL database run

- Linux operating system (OS) used on the server hardware
- Microsoft Windows OS installed on the Tek Panel display
- MoMA's network
- Audio soundtrack
- Image files
- Video files

To simplify our discussion, we will focus on four primary components that are part of *IWYTWM*.

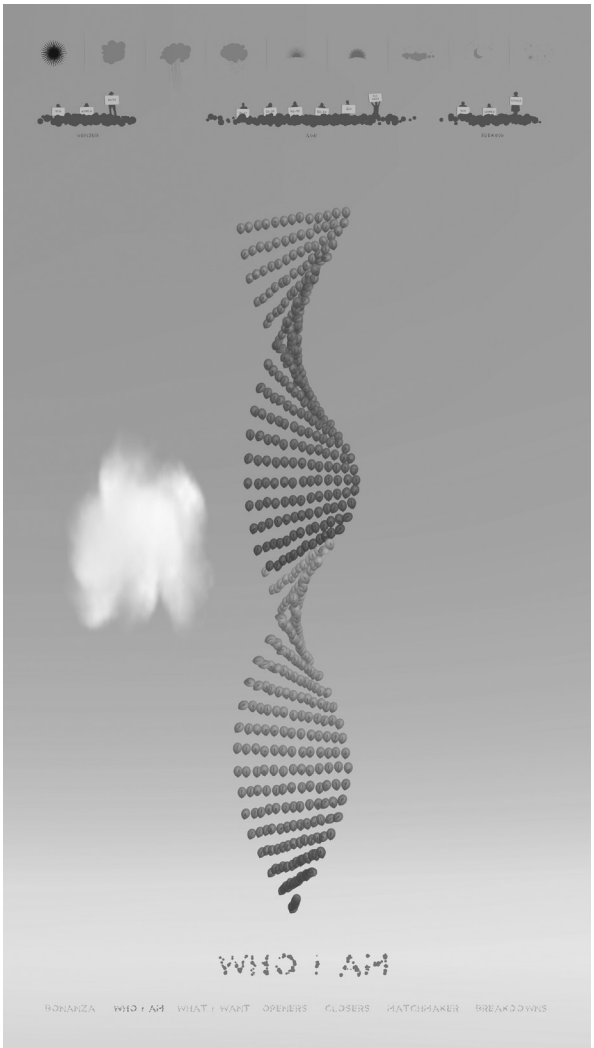

Fig. 1. Jonathan Harris and Sep Kamvar, *I Want You to Want Me*, 2008, (C++, OpenGL, Java, MySQL, touch-screen monitor), Museum of Modern Art, acc. no. SC527.2008. Image of touchscreen display for the "Who I Am" movement. Courtesy of Museum of Modern Art.

C++ custom application

The most visible component, and therefore considered the front-end of *IWYTWM*, is a C++ custom program written by Jonathan Harris. This component runs the user interactions and images that are seen on the custom touch screen.

Touchscreen display

A large, high-resolution touchscreen display is the other front-end component seen by the viewer. The display selected was a Tek Panel, model number UHR 560T, with 3840 x 2160 pixels, and a 6.5 ms response time. Tek Panel specializes in extremely high-resolution displays with integrated computers (Tek Panel 2011). Custom drivers were written to allow the screen to operate with

Fig. 2. Public interaction with *I Want You to Want Me*. Courtesy of Glenn Wharton.

the Windows operating system, according to the artists' specifications.

Java web crawler

Sep Kamvar wrote a Java-based web crawler. To prepare for the *Design of the Elastic Mind* exhibition, this crawler went out on the Internet to harvest dating profile data from over twenty websites. It is this data that appears in the balloons run by the C++ custom application.

IWYTWM dataset

The data harvested by the Java web crawler component is stored in a MySQL database. At the time that *IWYTWM* was commissioned, MySQL was an open-source database application.[3] It is worth highlighting here that the *IWYTWM* dataset is a unique component of *IWYTWM*, while the MySQL database application used in the exhibition was, at the time, open source—freely owned and accessible by anyone.

The four components mentioned above are each derived from distinct and in some cases highly specialized computer technologies. While some of the components have more significance than others, there is no clear hierarchical relationship between them (fig. 3). What exists instead is an interconnecting web of components, i.e., an information technology (IT) system. Unlike traditional, object –based artworks, *IWYTWM* must be perceived as an IT system. This system is an abstract and highly malleable set of relationships between the components.

The components have distinct roles in maintaining the entire system. Like a human body in which organs such as the heart and lungs are cared for in unique ways, each component is a member of a whole functioning system. Therefore, one cannot readily understand *IWYTWM* relationships using the hierarchical concepts applied to still or moving image art (e.g., video) such as *siblings* and *derivatives*.

To understand the malleability of an IT system, such as that employed in *IWYTWM*, consider the following example. Over ten years from now it is highly conceivable that the MySQL database application will no longer

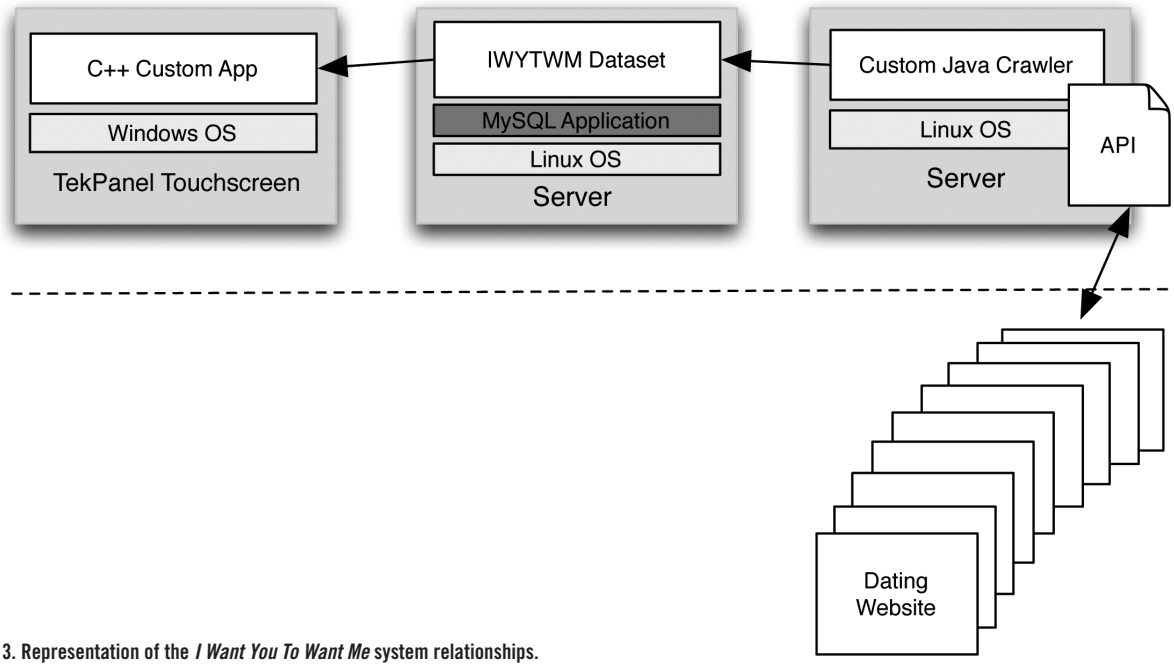

Fig. 3. Representation of the *I Want You To Want Me* system relationships.

be freely available for use by MoMA (especially in light of the fact the software corporate giant Oracle recently acquired Sun Microsystems and therefore also the copyright to MySQL) (Buckler 2009). While MoMA will still own the *IWYTWM* dataset, an alternate database application might prove to make more financial sense for the institution (Oracle licenses are known to be priced for enterprise corporate installations and could become prohibitive for a small implementation in an exhibit). The C++ application could still exist, and since MoMA has the source code it could ensure this component is able to work with an alternate database application or other data source chosen by the institution—maintaining the integrity of the abstract *IWYTWM* system even if one of the components has been modified.

MANIFEST CONSERVATION

Given the broad variation in technologies used for *IWYTWM*, it becomes fairly apparent that the preservation strategies employed will need to vary between the components. However, and perhaps most importantly, it is not just the variation in computer technologies that determines a component-based approach for conservation, since the artist themselves place different values on various components.

For example, the artists felt the Java web crawler was an integral part of understanding *IWYTWM*. However, they approved an exhibition of the work using a closed dataset (no longer networked), and retiring the functioning of the web crawler should it prove unfeasible to continue to crawl the dating sites (e.g., when the sites themselves no longer exist or the currently used data interface no longer exists). Consider Harris' (2008) take on the crawler:

> Certainly the easiest thing to do, and this would be very easy, would be to just stop crawling new data, but keep the whole database that already exists. And, you know, the piece has already been running for almost a year now, so, and *We Feel Fine*, our similar piece, has been running for five years now, and so, even if you were to stop off the new stream of data, you would still have an immense repository of historical data. And in a way, that becomes part of the work, this idea that this work could only have existed at this window of time, when dating sites were open, and people were using them, and the Internet was something we did on computers, and it wasn't yet a chip inside of our brains. It, in a way, it becomes like a little time capsule, like, this is a fingerprint of the world for these five years or these twenty years or whatever it ends up being. And I think that that's fine.

In contrast, Harris (2008) felt that for his C++ custom application

> The language that it's made in is very much a part of the piece. I don't think you can separate those things. I think you can separate the hardware that it runs on from the software, but I think the way the software is written is very much—there's even an artfulness in the code.

For Harris, maintaining the integrity of his application for as long as possible is integral to the work. Yet, not for the hardware: "The Tek Panel is not a requirement at all" (Harris 2008). Instead, he is more interested in maintaining the resolution standards, and the visitor experience achieved by the Tek Panel display.

Clearly it is not just the IT system relationships, nor the technology variations that challenge an object-based focus for conservation. Individual components may be assigned different values by the artists. This requires a

shift towards applying different conservation strategies to individual components, instead of a single strategy that is holistically applied to the work of art.

In addition, based on the artist statements highlighted above, several conservation levels begin to emerge for the artwork. A conservation level differs from a conservation strategy in that it is a recorded commitment an institution has made to an individual digital component for its long-term preservation. An institution chooses a conservation level for each digital component at the time it assumes preservation responsibility for a digital work in order to guide long-term decisions made by curators and conservators when applying conservation strategies. Selecting a conservation level is similar to determining whether a component is dedicated or non-dedicated to the work. However, digital components require a more nuanced interpretation. The level applied to a digital component can affect the types of systematic or routine preservation practices that are applied to it (e.g., whether to bulk migrate a deprecating file type to a new fully supported preservation-file type). Since a multi-component digital work of art would require differing conservation levels for the differing components, this leads us to argue for a multi-level conservation that includes a *manifest* conservation approach to renderability, bit preservation, and equipment storage.

The concept of manifest conservation, as it is applied in this article, is derived from the concept of renderability, a commonly understood concept in the digital preservation field. According to the OCLC/RLG Working Group on Preservation Metadata report, titled *Preservation Metadata and the OAIS Information Model*, "Renderability refers to the translation of the bit stream into a form that can be viewed by human users, or processed by computers" (OCLC/RLG 2002, 1). The report goes on to split renderability into two parts. First, a digital artifact should be transformed into a format that is compatible with hardware and software that can access it (if the original is not already in such a format) and then a person must also be able to be access and display the artifact as originally intended. David Holdsworth (2007, 8) sums the viewpoint nicely, "The key to doing things properly is to take a view of digital data as an abstract quantity, divorced from the medium upon which it is stored, but associated with information (technical meta-data—often including software) that permits ready access to its intellectual content."

While renderability provides a useful starting point, it is not entirely applicable to museum art collections. The primary distinction is that museums have a more vested interest in maintaining original formats of digital-based components and would not readily translate a digital file into an alternate format. While the library science field advocates translating a digital artifact into an archival format, if needed, after the institution assumes responsibility for it. For example, data files acquired by an institution might be translated to XML—regardless of the data's original format—so that the institution could standardize, and reliably ensure access to its collections data. This is because a reliance on original hardware and software is an inherent risk factor, as they will become obsolete. This is an inherent risk factor for art collecting institutions as well, but a risk that might be necessary to undertake for certain artworks.

Therefore, a new term is needed: manifest conservation. The implication is that the museum is dedicated to ensuring that the *experience* provided by a component can be replicated, as it was originally intended (just as with renderability). Unlike a library, a museum may defer data format translation until needed, or it may never decide to translate the digital file as the means for access. Consider the C++ custom application of *IWYTWM*. Given that Harris feels strongly that his code is significant, a manifest conservation level would be appropriate.

While a museum might not readily translate an original digital artifact that has manifest conservation level applied, as the technologies die that enable access to the

work (e.g., the Tek Panel display and the Microsoft OS installed on the touchscreen), a museum may employ various conservation strategies to maintain the viewer experience intended by the artist (DOCAM 2011a). To provide a tangible example, when the Tek Panel dies and there is no longer hardware available to run the original Microsoft OS, there is still a responsibility to render the C++ custom application. Given Harris's emphasis on the code itself, but lack of interest in the life of the hardware, an acceptable strategy might be to recompile the C++ for new hardware and a new OS—thereby maintaining the code base but still allowing the public to interact with the data as originally intended. Therefore, various preservation strategies can be applied throughout an artwork's lifespan (and most likely will be) to a component that has the manifest conservation level applied, including recompilation, emulation, and migration. However the conservation level would remain the same: the institution would ensure the *intended experience* is accessible.

While renderability is not an appropriate conservation level for the C++ custom application (based on the artist's statements), it may be appropriate for the *IWYTWM* dataset. Renderability would ensure access to the dating website data that has been harvested, and it would assume a commitment to a means to understand the data as originally intended—yet conservation of the original database schemas, and applicable technical standards are not necessary. So, while MoMA has assumed responsibility to conserve the dataset, it might not have a responsibility to conserve the MySQL database software (a particularly useful distinction in light of the fact that the Oracle Corp. has now acquired the MySQL codebase). This enables MoMA to transfer the data to new data store practices as they emerge in the future.

Another conservation level is bit preservation. Bit preservation is defined pretty much as it reads: a dedication to preserving the original bits of a digital artifact exactly as they originated. This should be understood as more than simple digital storage, as best practices include a dedication to ensuring the integrity of the digital artifact. The OCLC *Digital Archive Preservation Policy and Supporting Documentation* (2006, 4) states, ". . . activities for bit preservation include verifying the fixity of the objects and metadata and checking for viruses at the time of ingestion and annually on the anniversary of their ingest."

As previously stated, Kamvar does not require that access to *IWYTWM*'s Java web crawler be maintained in perpetuity, but he does want MoMA to maintain the code so that conservators, artists, and other interested communities are able to fully understand the artwork. Therefore, a bit preservation level could provide an appropriate solution for this component.

Still another conservation level that could be appropriately applied to *IWYTWM* is equipment storage. This should not be confused with the storage preservation strategy, which is most commonly defined as, "The principle generally applied in long-term preservation . . . to secure many duplicates of a given piece of equipment if its condition can be maintained in a storage environment when not in use" (DOCAM 2011b). While an equipment storage conservation level does involve storing a piece of equipment when not in use, the emphasis is on taking steps to extend the life of a computer component (e.g., using dust-free cool storage) because it is a valued artifact—but not because it is an intended means to conserve an artwork. The conservation level of media storage could be appropriately applied to the Tek Panel since the artists do not consider it part of the artwork, but it is still valuable because: (1) it was the original equipment the artwork was exhibited with; (2) it is an expensive artifact and so should be used again if feasible; and (3) it currently provides a means to access *IWYTWM*. However, what is important to note is that according to the artists, MoMA needn't assume a commitment to conserve it as a functioning component in perpetuity (which would be highly expensive and inherently unachievable).

JUSTIFICATION FOR A DIGITAL REPOSITORY

Complex artworks like *IWYTWM*, with varying and evolving technologies and multiple conservation levels, present complex demands for a museum. Conservators and IT specialists must

- maintain individual components at their appropriate conservation level;
- maintain appropriate access levels for the various components;
- record when and how computer code has been altered and recompiled to allow full renderability;
- accurately archive data formats and standards that are used to allow rendering of various components;
- record the complex "voices" that will be brought in to document and conserve the artworks; and
- record the highly complex and evolving IT system relationship that exists not only between the current components, but the replacement components, operating systems; compilers, and re-architected works in perpetuity.

Clearly these are resource intensive activities that provide strong justification for an IT system.

It is useful to highlight the differences between a dedicated digital repository used for long-term conservation of digital components and other IT systems used by collecting institutions, such as a collections management database, an internal file storage server, and a digital asset system. In 2002, the Consultative Committee for Space Data Systems (CCSDS), an international consortium of science institutions that develops data management standards, issued a Blue Book containing the *Reference Model for an Open Archival Information System (OAIS)*. The model proposed by CCSDS has since achieved ". . . the status of a de facto standard in digital preservation" (OCLC/RLG 2002, 1). The model is technology agnostic, in that it does not advocate any specific IT technology. Instead, it is a high-level outline that recommends not only how an IT system might support preservation issues for digital assets, but also the practices a collecting institution should follow in conjunction with an IT system to best ensure longevity of its digital assets. The mandatory responsibilities taken from the Blue Book have become a reference point for defining the purpose of a digital repository. A dedicated repository has a responsibility to

- Negotiate for and accept appropriate information from information producers
- Obtain sufficient control of the information provided to the level needed to ensure long-term preservation
- Determine, either by itself or in conjunction with other parties, which communities should become a "designated community" (CCSDS 2002, 1-10) and, therefore, should be able to understand the information provided
- Ensure that the information to be preserved is independently understandable to the designated community. In other words, the community should be able to understand the information without needing the assistance of the experts who produced the information.
- Follow documented policies and procedures which ensure that the information is preserved against all reasonable contingencies, and which enable the information to be disseminated as authenticated copies of the original, or as traceable to the original
- Make the preserved information available to the designated community (CCSD 2002)

In summary, a digital repository (the IT system and an institution's staff) has a responsibility to ensure it only accepts objects for which it can make reasonable assurance of longevity; that the intended viewer experience of these objects remains accessible and understandable; that access is highly controlled so that the digital integrity of the repository's file can be ensured; that access is controlled so that only those that should have access do;

and that all reasonable measures are taken to preserve a digital component and that a record of these measures is recorded in order to fully understand a current manifestation of an object.

While it is conceivable that an institution's collections management system (and supporting IT systems) could be extended to support a digital repository, it should not be assumed that a museum's collections management system (i.e., the technologies used and the institutional practices) could readily assume a digital repository's mandatory responsibilities. A full outline of the characteristics of a digital repository that strives to meet these mandated responsibilities, and how these differ from some other systems, is beyond the scope of this article. However, it is worth highlighting some differences at a high-level. The following characteristics are paraphrased from the report, *Trusted Digital Repositories: Attributes and Responsibilities* (RLG-OCLC 2002):

- *A digital repository is not just file storage:* A digital repository ensures integrity of its digital files, and therefore should not be understood simply as file storage. As mentioned earlier, preservation of digital components includes activities such as verifying the fixity of the digital files and metadata, checking for viruses at standardized checkpoints (such as at the time a file is acquired and as regularly scheduled maintenance. Records of these types of activities are meticulously captured.
- *A digital repository is like a vault:* As the report states, "A reliable digital repository: . . . obtains sufficient control of the information provided to support long-term preservation" (RLG-OCLC 2002, 21). Access is controlled not only for copyright or intellectual property rights reasons, but also to maintain the integrity of the digital files within a repository. For example, a digital repository would create a record any time a file is altered or copied and a traceable record back to the original would be maintained (including a record of the person or system that created the changes). Given this highly controlled access, it should not be confused with a collections management system that much of an institution's staff has access to.
- *Metadata is highly valuable, formal, standardized, and strictly applied:* The role of metadata—to ensure the integrity of its digital files, to control access, and to ensure its digital components remain accessible and understandable in the future—cannot be overstated. The metadata in a digital repository should not be understood as descriptive (e.g., name, title, date, description, etc.). While a digital repository could contain some descriptive metadata, more important to the purpose of a digital repository is the technical metadata that allows a digital file to be accessed in the future (e.g., an encoding standard is captured so that a future computer can understand how to interpret the computer bits within a file). Also important is the data that captures the relationships that exist between components so that a full IT system can be realized (e.g., the operating system needed to play a piece of code). Since without certain pieces of technical metadata an object could not be experienced as intended and the relationships between components are complex and ever evolving, the metadata that is captured in a digital repository needs to be formally and strictly captured and maintained.
- *A digital repository is not responsible for providing an object's viewing.* While a digital repository has a mandated responsibility to ensure its digital files remain accessible and understandable, it does not have a responsibility to provide a viewing experience. For example, a digital repository might store preservation quality videos, the appropriate codecs needed to access these videos, and possibly a copy of some software that can be

> used to play the video. However, the repository does not have a mandatory responsibility to have the codec and player installed to enable viewing access to this video. Other curatorial, collections management, and exhibition systems fulfill this role.

A COMPONENT-FOCUSED APPROACH: EXTENDING PREMIS

As already mentioned, the importance of the metadata within a digital repository cannot be overstated. For example, Caitlin Jones and Lizzie Muller (2008, 419) state in *Between Real and Ideal: Documenting Media Art*,

> Media artworks rarely exist as static, discrete and unique objects, but rather as collections of components, hardware and software, which together create time and process based experiences. Such works may change radically depending on the contextual conditions of their staging. Even the material components of such works are subject to rapid change due to technological obsolescence. Documentation is, therefore, increasingly important in media art, as it provides a continuing source of knowledge as to how a particular work manifests over time.

Adding to the importance of metadata is the fact that the IT architectural framework of any digital repository is its metadata schema. During the spring of 2008 to winter of 2009, the author, Barbra Mack, investigated several existing documentation practices and metadata schemas that were in use by the larger digital preservation field as well as the more narrow art collecting field (Mack 2008). This investigation was conducted as preliminary research to justify the need for a dedicated digital repository for MoMA. The schemas were evaluated with an eye towards museum needs, as well as to identify a data model that could support works employing an abstract IT system such as *IWYTWM* without creating obvious data capture limits.

There have been a number of efforts to standardize collections and preservation data and some well-known schemas already exist, including several based on the OAIS model. Some popular schemas include: METS (Metadata Encoding and Transmission Standard), PREMIS (Preservation Metadata: Implementation Strategies), MARC (Machine-Readable Cataloging), and Dublin Core. It should be noted that these schemas have originated primarily within data archives and the library science fields. While they provide significant insight for art collecting institutions, it should not be assumed they could be wholly applied to an art collecting institution's digital repository. On the other hand, there have also been several important projects that have led to varied art-specific data schemas, unique to each project, that do not identify themselves as OAIS based, including the Documentation and Conservation of Media Arts Heritage (DOCAM) Documentation Model (2011c); the Variable Media Questionnaire (2011); the Media Art Notation Model (MANS) (Rinehart 2011); and the International Network for the Conservation of Contemporary Art (INCCA) Artists' Archives Database (2011).

From this investigation it was clear that no single schema yet exists for the museum field that captures the full range of metadata needed to conserve computer component data. As Jones and Muller (2008, 418) state,

> traditional models of documentation are not well adapted to such works. Recent attempts to develop new models for documenting media art offer flexible paradigms, which focus on the processes of creation and exhibition, rather than on static objects. However, there is still an important gap around the documentation of the audience's experience of the work,

and ways to integrate experiential documentation with other information.

One element contributing to the documentation gap identified by Jones and Muller are data models that are often employed. Most existing schemas employ an object-based approach, in which metadata is hierarchically structured. They assume the existence of an originating object—which serves as the main entry and the parent of all other data entered. From this original object, layers of metadata are then applied, such as format information, technical specifications (e.g., hardware used, software, etc.), physical descriptions, time-based information (e.g., seconds and hours recorded), and so on. An object-based approach is problematic, because as already detailed earlier, multi-component digital works of art employ an abstract IT system that cannot be understood hierarchically.

This article does not attempt to formally critique any existing model. Each provides valuable insight, enabling the emerging field of digital art conservation to move towards establishing well-informed best practices for a digital repository specific to the conservation of computer art. Instead, the authors wish to highlight that regardless of the eventual data schema applied within an institution's digital repository, the data model should be component-focused rather than object focused in order to be sustainable for complex digital, multi-component works of art.

As part of MoMA's preliminary investigation, the PREMIS schema was identified as being particularly appropriate. PREMIS is both a data dictionary for preservation metadata and an international working group dedicated to developing the metadata schemas specifically for digital preservation. PREMIS shows particular strength for the following reasons:

- PREMIS is based on METS, a well-established metadata schema within the digital preservation field
- As an extension of METS, it is a data model based on the OAIS reference model for a digital repository
- The PREMIS data dictionary includes a granularity of semantic units (i.e., sublevels of metadata) for software and hardware attributes that did not exist at the time of the investigation in other common schemas
- A single digital component, which exists as a part of a multi-component work, is treated as an individual object within a PREMIS based repository. It can then be linked to other objects in various forms in order to create a cohesive "representation" of a work of art. This multi-directional linking approach provides a flexibility that is needed to support multi-component works of art that simply cannot be hierarchically organized under a single object. Note that the term "object" in the PREMIS model should not be confused with the museum world's use of the term. Within the PREMIS model, an object is more closely aligned with the term "component" and a "representation" is more closely aligned with the museum concept of an "object" (see below for more details).
- Treating each component as an individual entry in the system allows more extensibility to the metadata schema to enable support for wide variety of object types

While PREMIS has more metadata elements for software and hardware than many other schemas evaluated, these elements were not deep. In addition, PREMIS was not developed for art. So, as part of MoMA's preliminary investigation of documentation practices in the preservation field, MANS, DOCAM, and the Variable Media Questionnaire were carefully assessed. The PREMIS schema was extended to support computer-based art

conservation metadata needs by the author Barbra Mack, with significant input from Chris Lacinak and Glenn Wharton (Mack 2008a). This expansion of PREMIS was conducted as a means to investigate the feasibility of using it for conserving multi-component computer art. For this investigation, the PREMIS model was understood and applied as outlined in figures 4 and 5 and in the Appendix.

Using *IWYTWM* as an example, the Intellectual Entity (as listed in the Appendix) is the artwork. One representation of *IWYTWM* might be a collection of objects that make up the components that are inherent to the work (e.g., the original C++ custom component, custom Java web crawler, etc.). Another representation could be created to group the objects needed for a particular exhibition (e.g., C++ custom component *plus* the custom Java web crawler *plus* the Tek Panel high resolution monitor, etc.). While the C++ custom component could be grouped within both representations there would only be one entry for the original C++ custom application component in the repository, as well as its related metadata.

The benefit of each PREMIS object (i.e., the component) having individualized and highly granular semantic units (the granular object-level metadata), instead of apply-

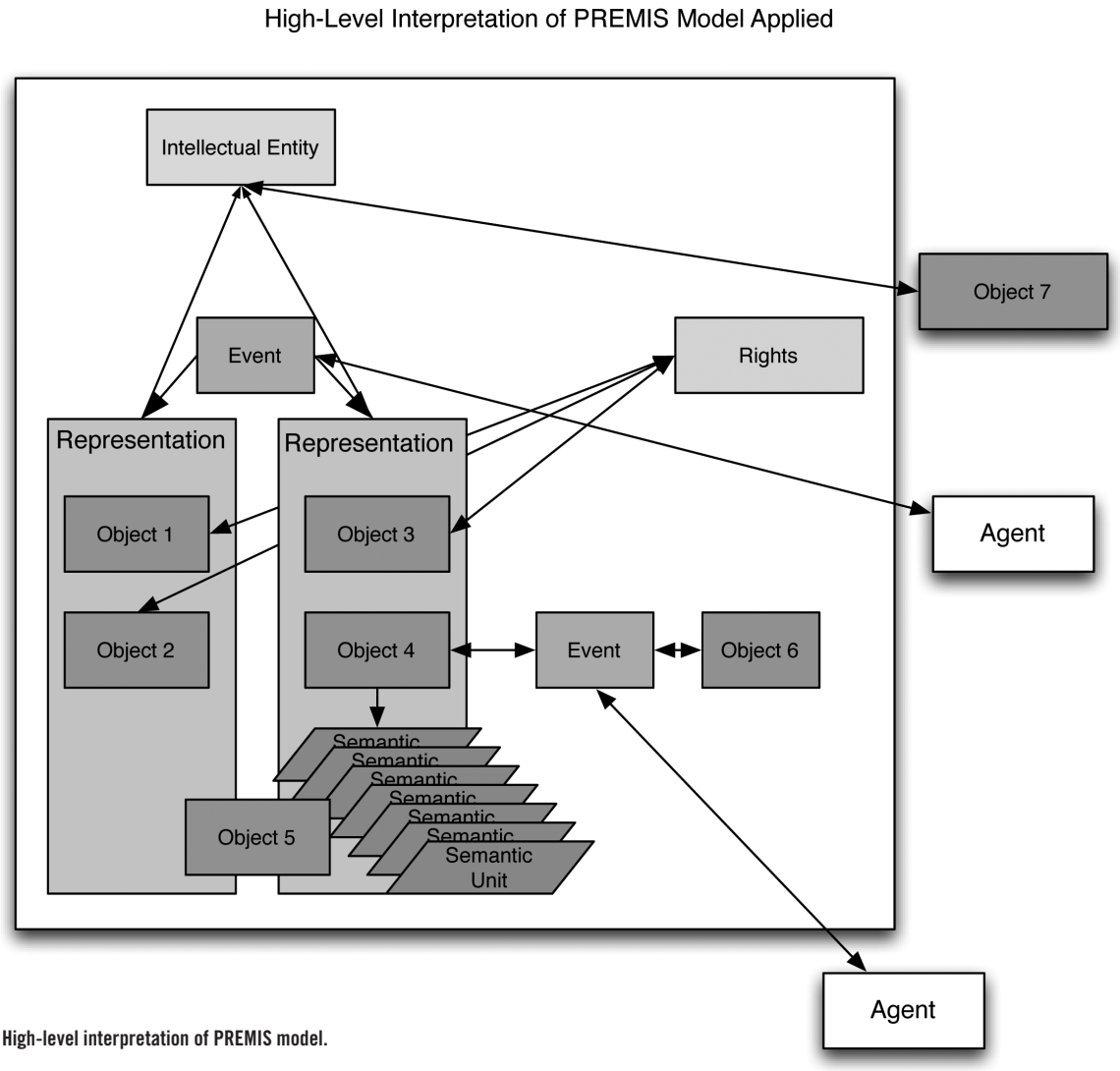

Fig. 4. High-level interpretation of PREMIS model.

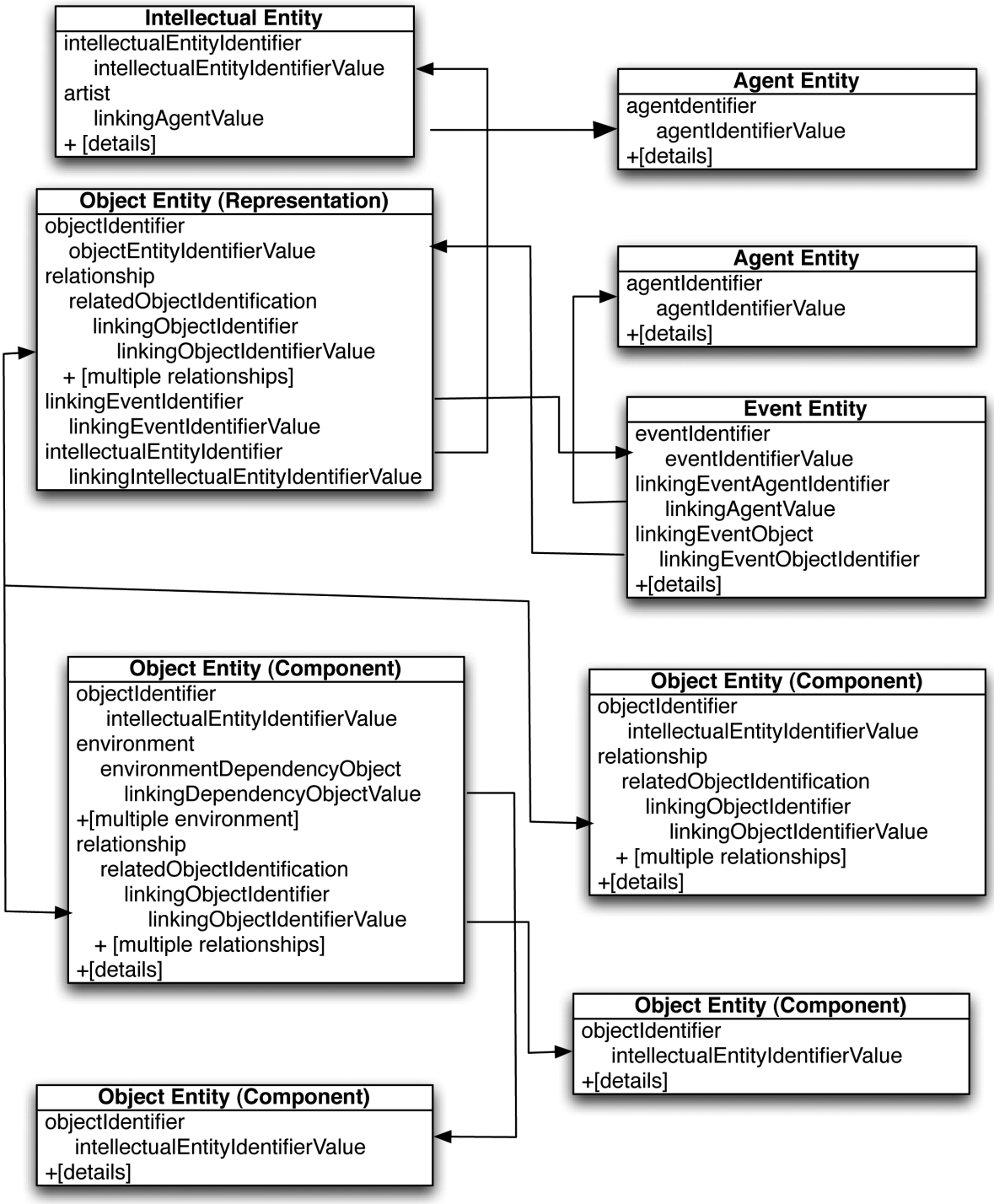

Fig. 5. Highly condensed view of the extended PREMIS data relationships.

ing metadata at the artwork level, cannot be overstated. The metadata for each component will be ever growing and more often than not will be highly detailed. Without a component-focused approach such as that used in PREMIS, the data dictionary would quickly break. For example, if a component-focused approach is not applied, how would the system hierarchically rank the C++, Java, MySQL data, or custom hardware? While a user might multi-select all of the formats that apply to an artwork, forty-plus years into the future, the Tek Panel for *IWYTWM* may stop working and the C++ custom application may need to be recompiled for an exhibition. Under an object-based approach, it would be difficult to readily and accurately answer questions about a single component such as

- which of the digital components is C++;
- which version of C++ was it;
- what hardware was the C++ custom originally run on and what changes would need to be made to the code so a new processor can interpret the application;
- how many times has the application been recompiled in the past;
- what code changes were made to the code as a preservation strategy;
- who made the changes;
- how does this application interact with the other components in the artwork; what is its intended experience;
- do C++ compilers still exist for current hardware and if not when and how was the original file translated; and
- what types of manuals, technical standards, and other technical documentation exist to assist in understanding and accessing the component?

And these are the questions for just one component. There are over eleven primary components of *IWYTWM*, and answering questions for each would require detailed and responsibly captured metadata. The preservation field is already aware of the level of metadata needed to responsibly conserve digital video, audio, and digital image objects. Consider that for a multi-component digital work of art, this level of data needs to be captured for *each* component.

Similar questions could be asked about the combination of components that constitute the artwork's abstract IT system. Including,

- which components are the originals;
- which component combination is considered the current manifestation of the artwork;
- which components were used in Exhibition A vs. Exhibition B;
- which components were replaced by derivatives or migrations of an original component and therefore constitute a new representation of an artwork?

With a component-focused model such as used by PREMIS, these types of data complexities are more readily recorded and resolved. The different combinations of components that equal differing manifestations of an artwork are simply captured as individual representations of the artwork. And since each PREMIS object (i.e., the component) can have its own highly granular and specialized metadata, the data captured can evolve elegantly for each component. In addition, individualized data schema extensions can be developed for different media types (e.g., audio, video, custom software, hardware, etc.) and be applied at the component level without breaking the original data model. This allows the digital repository to grow as new technologies and more refined data practices emerge without having to completely overhaul the IT system's architecture.

It should be noted that the PREMIS metadata schema does not currently go deep enough to capture all of the data needed to adequately conserve hardware and software. For instance, the museum model will need to

extend the schema to add metadata elements for serial numbers, version numbers, model numbers, etc. Various other metadata needs and standards need to be developed in this area of technical detail much further. In addition, a fully developed data dictionary for art collecting institutions would also need significant development in metadata elements specific to the artistic voice and any aesthetic characteristics that influence conservation practices.

While PREMIS might not prove to be the actual schema applied in the final implementation of MoMA's digital conservation repository, this preliminary investigation did highlight that a component-focused model will prove to be much more malleable than the object-based approach used in many previously developed schemas. A multi-component approach is needed to ensure that a digital repository for museum collections will meet the needs for long-enough duration in order to justify the institutional and financial commitment needed to build it.

CONCLUSION

In this article we demonstrated the need for a multi-component approach to building a managed conservation repository for digital artworks in museums. A number of useful models and schemas currently exist that can be used as building blocks for such a museum repository. OAIS and PREMIS from the library and archive world, and MANS, DOCAM, and the Variable Media Questionnaire from the museum world provide a useful foundation. We hope that this research will serve conservators, IT professionals, and others who are charged with managing information about digital artworks and that the analysis of *I Want You to Want Me* by Jonathan Harris and Sep Kamvar will serve as a reminder of the importance of honoring values assigned to underlying components by the artists themselves.

ACKNOWLEDGEMENTS

The authors wish to thank Jonathan Harris and Sep Kamvar for their assistance in this research—and for providing the world with such an intriguing work of art. We thank the IT staff, the Media Working Group at MoMA, and the Conservation of Computer-Based Art Working Group at New York University for the research and support that paralleled this project. We are indebted to Kara Van Malssen and Deena Engel for allowing us to use findings from their research on MoMA's digital collections for this article. In addition, we express our gratitude to Jim Coddington for reviewing the text.

APPENDIX

PRELIMINARY EXTENDED PREMIS DATA DICTIONARY

Intellectual Entity

- Intellectual Entity
  - intellectualEntityIdentifer
    - intellectualEntityIdentifierType
    - intellectualEntityIdentifierValue
  - title
  - artist
    - linkingAgentIdentifier
      - linkingAgentIdentifierType
      - linkingAgentIdentifierValue
    - linkingAgentName
    - linkingAgentType

Object Entity

- Object Entity
  - objectIdentifier
    - objectIdentifierType
    - objectIdentifierValue
  - originalName
  - preservationLevel
  - use
  - objectCategory
    - objectSubCategory
  - storage
    - contentLocation
      - contentLocationType
      - contentLocationValue
  - objectCharacteristics
    - installationVisible
    - installationInstructions
    - SignificantCharacteristics
    - Customizations
    - specTechnicalConsiderations
  - digitalFileCharacteristics
    - fixity
      - messageDigestAlgorithm
      - messageDigest
      - messageDigestOriginator
    - size
      - fileSizeUnits
      - fileSizeValue
    - format
      - formatDesignation
        - formatName
        - formatVersion

- formatRegistry
  - formatRegistryName
  - formatRegistryKey
  - formatRegistryRole
- fileExtension

- compositionLevel
- digitalTranslation
  - compressionApplied
  - encodingApplied
  - compiled
- sourceOwned
- commericalAppCharacteristics
  - applicationName
  - applicationVersion
  - commercialAppmanufacturer
    - linkingAgentIdentifierType
    - linkingAgentIdentifierValue
    - linkingAgentIdentifierName
  - ProductID
  - purchaseDate
- CustomAppCharacteristics
  - languageName
  - lanaguageVersion
  - customAppManufacturer
    - linkingAgentIdentifierType
    - linkingAgentIdentifierValue
    - linkingAgentIdentifierName

- hardwareCharacteristics
  - hardwareManufacturer
    - linkingAgentIdentifier
      - linkingAgentIdentifierType
      - linkingAgentIdentifierValue
      - linkingAgentIdentifierName
  - model
    - modelName
    - modelNumber
  - serialNumber
  - purchaseDate
  - input
    - inputType
    - inputQuantity
  - monitor
    - monitorType
    - aspectRatio
    - resolution
    - dotPitch
    - displayArea
    - responseTime
    - refreshRate

APPENDIX (CONT)

PRELIMINARY EXTENDED PREMIS DATA DICTIONARY

- viewingAngle
  - viewingAngleTheta
  - viewingAnglePhi
- luminance
- contrastRatio

▪ memory
- memoryQuantity
- memoryType
- memoryCapacity
  - capacityUnits
  - capacityValue
- cache
  - cacheCapacity
    - capacityUnits
    - capacityValue
  - cacheType
- dataTransferRate
  - dataTransferUnits
  - dataTransferValue
- processingSpeed
  - processingSpeedUnits
  - processingSpeedValue
- integerRange

▪ processor
- integerRange
- processingSpeed
  - processingSpeedUnits
  - processingSpeedValue
- cache
  - cacheCapacity
    - capacityUnits
    - capacityValue
  - cacheType

o objectPhysicalCharacteristics

▪ color

▪ material

▪ physicalMeasurements
- depth
  - depthUnits
  - depthValue
- width
  - widthUnits
  - widthValue
- length
  - lengthUnits
  - lengthValue
- diameter
  - diameterUnits
  - diameterValue

- audioCharacteristics
  - face
    - faceIdentifier
      - faceIdentifierType
      - faceIdentifierValue
    - faceLabel
    - region
      - regionID
      - regionLabel
      - noiseReduction
      - Equalization
      - speedAdjustment
      - playBackSpeed
        - playbackSpeedUnits
        - playbackSpeedValue
      - stream
        - streamed
        - streamLabel
      - channel
        - channelNumber
        - channelMapLocation
    - playbackEqualization
    - recordedDirectionsValue
    - potentialPlaybackTime
      - playbackTimeUnits
      - playbackTimeValue
  - analogTapeCharacteristics
    - backcoating
    - oxideType
    - screwsOnCassette
    - analogTapeLength
      - unwoundTapeLength
        - unwoundTapeLengthUnits
        - unwoundTapeLengthValue
      - tapeWidth
        - tapeWidthUnits
        - tapeWidthValue
      - tapeThickness
        - tapeThicknessUnits
        - tapeThicknessValue
- creationCharacteristics
  - dateCreated
  - createdBy
    - linkingAgentIdentifier
      - linkingAgentIdentifierType
      - linkingAgentIdentifierValue
    - creatingObject
      - linkingDependencyObjectIdentifier
        - linkingDependencyObjectIdentifierType

APPENDIX (CONT)

PRELIMINARY EXTENDED PREMIS DATA DICTIONARY

- linkingDependecnyObjectIdentifierValue
- environment
  - environmentPurpose
  - environmentNote
  - environmentDependencyObject
    - linkingDependencyObject
      - linkingDependencyObjectType
      - linkingDependencyObjectValue
- relationship
  - relationshipType
  - relationshipSubType
  - relatedObjectIdentification
    - linkingObjectIdentifier
      - linkingObjectIdentifierType
      - linkingObjectIdentifierValue
  - relatedEventIdentification
    - linkingEventIdentifier
      - linkingEventIdentifierType
      - linkingEventIdentifierValue
- linkingEventIdentifier
  - linkingEventIdentifierType
  - linkingEventIdentifierValue
- linkingIntellectualEntityIdentifier
  - linkingIntellectualEntityIdentifierType
  - linkingIntellectualEntityIdentifierValue
- linkingDocumentIdentifier
  - linkingObjectIdentifier
    - linkingObjectIdentifierType
    - linkingObjectIdentifierValue

Event Entity

- eventIdentifier
  - eventIdentifierType
  - eventIdentifierValue
- eventType
- eventDateTime
- eventDetail
  - eventDetailNotes
  - audioSettings
  - videoSettings
- eventOutcomeInformation
  - eventOutcome
  - eventOutcomeDetail
- linkingAgentIdentifier
  - linkingAgentIdentifierType
  - linkingAgentIdentifierValue
  - linkingAgentIdentifierRole
- linkingEventObject

    - linkingEventObjectIdentifier
        - linkingEventObjectIdentifierType
        - linkingEventObjectIdentifierValue
    - linkingEventObjectCategory
- linkingDocumentIdentifier
    - linkingDocumenttIdentifier
        - linkingDocumenttIdentifierType
        - linkingDocumentIdentifierValue
    - linkingDocumentDetails

Agent Entity

- agentIdentifier
    - agentIdentifierType
    - agentIdentifierValue
- agentName
    - Name1
    - Name2
- agentType
- agentDetail
- agentAddress
    - buildingIdentification
    - streetAddress
    - CityLocality
    - StateProvence
    - Country
    - PostalCode
- agentTelephone
    - TelephoneNumber
    - TelephoneType
- agentEmail
- linkingAgent
    - linkingAgentIdentifier
        - agentIdentifierType
        - agentIdentifierValue
    - linkingAgentName
    - linkingAgentType

REFERENCES

Buckler, C. 2009. Oracle publicly commits to MySQL. SitePoint Pty. Ltd. www.sitepoint.com/oracle-mysql-commitment (accessed 06/10/11).

CCSDS (Consultative Committee for Space Data Systems). 2002. Reference model for an Open Archival Information System (OAIS). Washington, D.C.: CCSDS Secretariat. http://public.ccsds.org/publications/archive/650x0b1.PDF (accessed 10/01/11).

DOCAM. 2011a. Conservation guide: a preservation guide for technology-based artworks. Documentation and Conservation of Media Arts Heritage. www.docam.ca/en/about-cons.html (accessed 06/10/11).

DOCAM. 2011b. Conservation guide: a preservation guide for technology-based artworks. 2.1.4 Storage. Documentation and Conservation of Media Arts Heritage. www.docam.ca/en/21-conservation-strategies/214-storage.html (accessed 06/10/11).

DOCAM. 2011c. DOCAM documentation model. Documentation and Conservation of Media Arts Heritage. www.docam.ca/en/documentation-model.html (accessed 06/10/11).

Engel, D. 2010. The Museum of Modern Art computer based collections risk assessment. Museum of Modern Art, New York, NY.

Forging the Future. 2011. Variable media questionnaire. 3rd generation beta. http://variablemediaquestionnaire.net/# (accessed 10/03/11).

Harris, J. and S. Kamvar. 2008. I want you to want me. http://iwantyoutowantme.org (accessed 06/10/11).

Holdsworth, D. 2007. Preservation strategies for digital libraries. DCC curation reference manual. HATII, University of Glasgow. www.dcc.ac.uk/resources/curation-reference-manual/completed-chapters/preservation-strategies (accessed 10/01/11).

INCCA. 2011. INCCA database for artists' archives. International Network for the Conservation of Contemporary Art. www.incca.org/idaa (accessed 06/10/11).

Jones, C., and L. Muller. 2008. Between real and ideal: Documenting media art. *Leonardo* 41 (4): 418–419.

Mack, B. 2008a. MoMA digital collections conservation repository phase 1 design documentation. Museum of Modern Art.

Mack, B. 2008b. Preliminary survey of digital artifact conservation practices. Museum of Modern Art.

MoMA. 2008. Design and the elastic mind. Museum of Modern Art. www.moma.org/interactives/exhibitions/2008/elasticmind (accessed 06/10/11).

Harris, J. 2008. Personal communication. November 17. Museum of Modern Art.

Kamvar S. 2008. Personal communication. November 21. Museum of Modern Art.

Kamvar S. 2010. Personal communication with Deena Engel and Mona Jimenez. April 2. New York University.

OCLC/RLG Working Group on Preservation Metadata, OCLC, and Research Libraries Group. 2002. Preservation metadata and the OAIS Information Model: a metadata framework to support the preservation of digital objects. Dublin, Ohio: OCLC/RLG Working Group on Preservation Metadata. Also available at www.oclc.org/research/activities/past/orprojects/pmwg/pm_framework.pdf (accessed 10/01/11).

OCLC. 2006. OCLC digital archive preservation policy and supporting documentation. Online Computer Library Center, Inc. http://www.oclc.org/support/documentation/digitalarchive/preservationpolicy.pdf (accessed 06/10/11).

Rinehart, R. 2011. A system of formal notation for scoring works of digital and variable media art. University of California, Berkeley. www.bampfa.berkeley.edu/about/formalnotation.pdf (accessed 10/03/11).

RLG-OCLC. 2002. Trusted digital repositories: Attributes and responsibilities. Research Libraries Group. www.oclc.org/research/activities/past/rlg/trustedrep/repositories.pdf (accessed 06/10/11)

Tek Panel. 2011. Tek Panel, Inc. www.tekpanel.com (accessed 06/10/11).

Barbra Mack
Creative Technologist
Brooklyn, NY
b.l.mack@blmak.com

Glenn Wharton
Conservator
Museum of Modern Art
11 West 53rd Street
New York, NY 10019
glenn_wharton@moma.org

Papers presented in *The Electronic Media Review* have not undergone peer review.