

Doing the DAM: Digital Asset Management at the Metropolitan Museum of Art

by Shyam Oberoi

Digital Images in Museums

Computer science has coined the term *GIGO* – garbage in, garbage out – which is another way of saying that a system is only as good as the information in it. A corollary to GIGO may be “a system is only as good as the people who advocate, develop and run it.” When implementing a digital asset management system (DAMS) – which allows for the storage of and access to a seemingly limitless quantity of image, sound and video files – the need for data integrity and development acumen can be particularly acute, especially if the intention is to provide museum-wide access to the DAMS and if the DAMS must sync with other systems already in place. It is imperative that the digital asset repository and its rules be clearly defined, intelligently organized and understood by those who will manage and use the system.

The Metropolitan Museum of Art (the Met) has a long history of producing, using and sharing images: from black and white negatives to color transparencies and, most recently, digital images. The Met Images project, which began officially in the fall of 2005, but which was discussed and planned long before, was formed so that the Museum could more systematically create, store, manage, catalogue and distribute digital images. The goals of the project were two-fold:

1. To support the museum’s core mission objectives to research, document and educate through an essential investment in the museum’s assets and infrastructure.

Shyam Oberoi is the manager of Met Images. Before joining the Metropolitan Museum of Art, he worked for a number of non-profit, corporate and governmental organizations, including the New York Bar Association, VA Hospital, MCI, Lehman Brothers and Accenture. He can be reached by email at shyam.oberoi@metmuseum.org

2. To strengthen the quality and quantity of available images and their cataloguing so that images could be quickly located and processed for distribution and licensing to both internal and external customers.

This project, one of the most expensive non-construction projects ever undertaken by the Met, involved close work with staff in our photo studio (who produce the images), in the image library (who distribute and license the images), in our various curatorial departments (who research and record information about our collection) and in our IT department (who ensure that all systems used by these staff function smoothly).

In cooking, preparation takes the longest time and is the least rewarding, yet it is essential in order for the process to go smoothly. The same principle applies to system development. In implementing a DAMS at the Met, we began with two main ingredients: records in our collections management system for objects in the museum’s permanent collection and several thousand digital images of and related to those objects. This article examines the unique challenges presented in preparing each of these ingredients to be ingested into a new DAMS and underscores the notion that, no matter how much you prepare, developing a complex mechanism like a DAMS is a fluid and open-ended process that brings issues of sharing and access to the fore.

Object records. The Met has used The Museum System (TMS) as a collections management system since 1995 and began implementing MediaBin, a digital asset management system, in 2005 [1]. TMS is the museum’s central repository for information about objects in its collection, including information about an object’s exhibition, publication, valuation and conservation history. MediaBin is intended to be the museum’s central repository for images of and related to those objects (among other types of images). We needed a way of both porting object records from TMS into

MediaBin and keeping information about objects within these two systems in sync, in order for images to be found in MediaBin and to forestall user confusion over data shared by the two systems. We quickly determined that we would not be able to bring all information related to an object into MediaBin because MediaBin's metadata structure does not lend itself well to complex data relationships, as does TMS. Instead, we limited the object data brought into MediaBin from TMS to "tombstone" data, that is, the types of information you would typically see on a wall label in a museum gallery (e.g., artist name, birth and death dates, object title and date). In addition, we determined that some object information would need to be shoehorned into single fields in MediaBin. TMS has, for example, a repeating artist field, allowing the museum to record any number of "child" artists for a single "parent" work. MediaBin, on the other hand, prescribes use of a single artist field, forcing the Met to concatenate artist name data in MediaBin for works with multiple artists.

Digital images. Given the enormous number and diverse types of images produced by the Met, we resolved to limit our initial ingest of images into MediaBin to the entire digital production of our photo studio: over 200,000 digital files representing official images of objects in the Met's collection, captured over a decade of digital photography and scanning, and occupying more than 4 terabytes of disc space. Administrative information for these images, such as photographer name and rights-related information, had previously been stored in a number of legacy systems, ancillary to TMS, including a FileMaker database which recorded the contents of some of the DVDs used previously to store digital images, a ColdFusion application for ordering photography and a separate database for tracking black and white negatives. The photo studio's images represent only a small percentage of the total digital resources produced throughout the museum. Not included in the initial ingest, for example, were images from our conservation departments and photographs of excavations taken by museum staff working at archeological sites. While all images will eventually be ingested into the DAMS, we decided to deal with those images at a later date, after understanding more precisely how museum staff actually use the system.

Throughout this initial planning process, we had to make decisions about technical and organizational matters. Our decisions reflected how we imagined information in the DAMS should be organized and how, given our available resources, we might best be able to transition that information from existing locations, such as a collections management system and legacy systems. We found that our decisions related to organizing information often differed from those made by some of our sister institutions when implementing a DAMS. For example, object records, along with their associated image files, can be organized in MediaBin using a hierarchical tree-like structure similar to a file system, with parent folders (containers) and child subfolders (sub-containers), either of which might include some combination of object records and image files. Some of our sister institutions decided to make separate parent folders for every object record and place associated image files into child subfolders within that parent folder. However, given the size of the collection at the Met, we determined that such a granular approach would be too unwieldy. Instead, we instituted a hierarchy that mimicked the organizational structure of the museum, with separate parent folders for each of our 18 curatorial departments and then with child subfolders for the various types of assets (that is, digital images, analog media such as color transparencies and black and white negatives, and object "placeholder" records from TMS).

Making Connections

Making the connection between different sets of data – object information, images of those objects and image information – was one of our greatest challenges in implementing a digital asset management system. Information about objects in the Met's collection is maintained in TMS, our collections management system, and information related to images has been stored in a number of legacy, ancillary systems. Not surprisingly, the specific metadata fields as well as the format and order of these fields varied from system to system. In addition, the quality and integrity of these data varied and needed to be reconciled from these various systems before bringing the data into MediaBin. For example, object information needed to be pulled from TMS, while image information such as "photographer" and descriptions

of the image views (such as whether it was overall or detail) was recorded in Photo Studio Workhorse, our legacy ColdFusion application for ordering photographs. Finally, the contents of CDs and DVDs created in the photo studio before the introduction of the Filemaker database were recorded in Excel spreadsheets.

Since digital images do not yet exist for every analog transparency and negative in the museum, we decided to bring records about these types of assets into MediaBin as well, in order to indicate to our users that these images exist and can be scanned, a considerably less expensive option than ordering new photography. Information about negatives was stored in a separate SQL database, while information about our boxes of color transparencies had been transcribed into additional Excel spreadsheets.

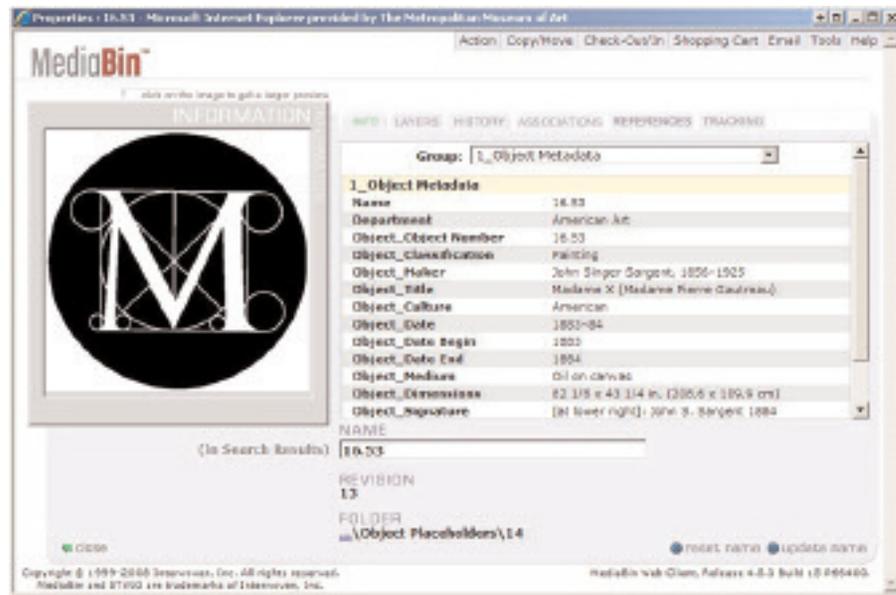
Gathering, organizing and understanding these different sets of data required coordination with multiple stakeholders within the museum. Curatorial departments contribute and maintain their own object information within TMS, information that is constantly changing as a result of ongoing research and new acquisitions. In general, direct access to object information in TMS is limited to staff in the curatorial department to which the object belongs. Information about image rights and licensing restrictions is maintained by the image library, the department responsible for the distribution of images both within and outside of the museum. Interestingly, before the introduction of MediaBin, this department did not have an automated way of recording or retrieving this information. Finally, the photo studio, with its team of photographers, scanners and post-production staff, creates the museum's official images of objects in the Met's collection.

The first step in readying our TMS object information for porting into MediaBin was to create an extract datafile. This datafile contains a single, flattened record for each TMS object record limited to the tombstone data for an object. The datafile is refreshed nightly with the latest data about objects in TMS and then picked up in MediaBin. In addition, a checksum value is generated in the datafile for each TMS object record and then compared to that record's value from the previous night's data sync. Any records with different checksum are updated, and any new TMS object records such as those for new acquisitions are inserted into the datafile for

subsequent refresh in MediaBin. This process proved to be relatively straightforward since TMS and MediaBin run on SQL Server, and we have a similar extract scheme in place for other systems within the museum that feed off TMS object information.

The initial preparation of our digital images was a bit more cumbersome. The photo studio's discs – all 4,000 of them – first needed to be loaded onto the network for ingestion into MediaBin. Once the image files were copied, we began the process of normalizing the spreadsheets describing the disc contents and evaluating the usefulness of the digital files. For example, duplicate image files were removed, and legacy file formats were converted to TIFF. A significant amount of work needed to be done first to massage data in the spreadsheets because data was often missing and in all cases incomplete and then in order to connect the image files to the TMS object records in our ingest datafile. Once in MediaBin, image files related to objects are connected to the correct object data via a bi-directional association (or relationship). Figure 1

FIGURE 1. TMS object record represented in MediaBin



OBEROI, continued

FIGURE 2. Images associated with that object record



and Figure 2 illustrate associations made between the object record for accession number 16.53 (the painting, *Madame X*, by John Singer Sargent) and several images of that object. In other words, 16.53 is depicted in the image file named DT91.tif, and DT91.tif is an image of the object 16.53. Once image file and object record are linked, the tombstone data extracted from TMS is copied from the object record onto the image record in MediaBin.

Try, Try Again

Our initial ingest datafile contained a record for every object represented in our collections management system, every official digital image produced by our photo studio and all records for color transparencies and black and white negatives pulled from ancillary systems. Rather than attempting to process these all at once, we resolved to ingest this information in three main stages:

1. **Object information:** all non-image records (that is, the TMS object

records and “placeholders” for analog material) were read from the datafile and ingested into MediaBin, divided into manageable batches of 50,000 records. A polling process was written to monitor the ingest, so that as soon as one batch finished the next could begin immediately.

2. **Images:** After the photo studio’s over 200,000 images were copied from the network into MediaBin, the same script was run again, this time processing only image records in the datafile. This process essentially revised each image in the MediaBin repository with information previously held in the spreadsheets and the ancillary systems.
3. **Associations:** Finally, once object and image records were ingested, a script was run to programmatically establish the bi-directional association between the two.

The process summarized above obviously represents a gross oversimplification of what actually took place. And while it was during this stage that the benefits of our planning and preparation were made apparent, it was also at this point that the things we had overlooked came to light.

Doing the Work

As noted, a system is only as good as the people who advocate, develop and run it. (We have all had experiences where a sluggish machine or an ailing database has been miraculously resurrected by some guru.) Although the Met retained consultants at the start of this process, staff within the museum eventually did almost all of the work outlined above. There is a strong temptation these days among many organizations to try to outsource as much information technology work as possible. For museums, given the challenges in funding and retaining resources, this temptation may be especially great. However, I think I can state with a fair amount of certainty that if the Met had relied wholly on consultants to implement its DAMS, we wouldn’t have one today.

When considering outsourcing, organizations need to make the distinction between maintenance and development. Maintenance is a natural candidate for outsourcing: the system is in place, is documented and is functioning in a regular and repetitive fashion. Development, however, is a

fluid process and requires constant discussion and reevaluation. In the case of implementing a DAMS – where the original source information for objects and images may be coming from any number of disparate systems and where the destination repository allows for unlimited variation in both organization and metadata definitions – the development effort can be particularly open-ended. No matter how much you prepare, at some point you will discover that something has been overlooked, that one of your assumptions is wrong, that some of your specifications have changed. At the Met, unfortunately, we made these discoveries more than a few times during our implementation. Trying to communicate these multiple changes to consultants became so complicated and the turn-around time on fixes so long, that we ended up bringing all of the work back in house.

It is impossible to overstate how valuable this in-house operational knowledge of the DAMS has been. Feedback from users and day-to-day observation of how the system is actually being used have led to modifications of our processes and customizations to the front-end MediaBin Web client. This kind of rapid, flexible development would be impossible had museum staff responsible for running and supporting the system not gotten their hands dirty during the implementation process. It would also not be possible if MediaBin did not have its own robust API.

Different organizations will have different ways of meeting development challenges. The resources which the Met, given its size, is able to devote to these issues will probably seem immoderate to some other cultural institutions. On the other hand, certain private sector corporations spend more in a single month on free food for their employees than the Met has spent during this entire multi-year process. In either case, our experience suggests that there are no shortcuts, no easy answers and no way to escape the fact that a DAMS is a complex mechanism which, like any enterprise-level application, requires a significant amount of supervision and technical expertise and touches on a range of different information technology and management skill sets, including database administration, web application development, network administration and storage and backup strategies.

Information Wants to Be Free?

In certain institutions today, the argument one most commonly hears against a free and open exchange of information is that content cannot be shared since it may end up in the hands of someone who does not understand it. It is an argument that was advanced more than two thousand years ago:

I cannot help feeling, Phaedrus, that writing is unfortunately like painting; for the creations of the painter have the attitude of life, and yet if you ask them a question they preserve a solemn silence. And the same may be said of speeches. You would imagine that they had intelligence, but if you want to know anything and put a question to one of them, the speaker always gives one unvarying answer. And when they have been once written down they are tumbled about anywhere among those who may or may not understand them, and know not to whom they should reply, to whom not: and, if they are maltreated or abused, they have no parent to protect them; and they cannot protect or defend themselves. [2]

Socrates was, of course, speaking about what was at the time a recent invention: the written word. Two thousand years later we find ourselves at a point not so dissimilar to that of the ancient Greeks, thus confronted with new technologies that are radically altering the ways in which information is disseminated. And yet, like Socrates, many organizations today don't yet understand that the old way of controlling information is essentially over. In museums, this mindset may be especially strong since in some ways the core premises of museums and technology are antithetical. On the one hand you have institutions whose entire raison d'être is the conservation and understanding of the past; on the other, you have a discipline almost rapaciously dedicated to innovation, to replacing the old with the new.

The introduction of an enterprise-level digital asset management system at the Met has brought issues of sharing and access to the fore. Not only have images from disparate sources – negatives, transparencies, digital images – been centralized, but core object information, which previously had been limited to staff of the curatorial department, is now accessible

OBEROI, continued

throughout the museum via the DAMS. In the coming months Met staff beyond the photo studio will begin contributing their own local image collections to the digital asset management system: x-rays of object conservation, maps designed for education and special exhibitions, even portrait photography of events within the museum. We have only just begun confronting the possibilities and challenges inherent with these new modes of access to images and information. ■

Resources Cited in the Article

- [1] Key applications discussed in this article are The Museum System (TMS), a collections management system produced by Gallery Systems; and MediaBin, a digital asset management system produced by Interwoven.
- [2] Jowett, Benjamin (1871). *The dialogues of Plato* (Vol. 1). Oxford: Oxford University Press.

New ASIST Titles from Information Today, Inc.

Call for our current catalog or visit www.infotoday.com

**Information and Emotion:
The Emergent Affective
Paradigm in Information
Behavior Research
and Theory**

Edited by Diane Nahl and Dania Bilal
ASIST member price \$47.60

ARIST 41

Edited by Blaise Cronin
ASIST member price \$99.95



**Covert and Overt: Recollecting and
Connecting Intelligence Service and
Information Science**

Edited by Robert V. Williams
and Ben-Ami Lipetz
ASIST member price \$39.60

Theories of Information Behavior

Edited by Karen E. Fisher,
Sanda Erdelez, and
Lynne (E. F.) McKechnie
ASIST member price \$39.60

Note: Prices do not include shipping and handling.



Information Today, Inc.

143 Old Marlton Pike • Medford, NJ 08055 • Phone: (800) 300-9868 or (609) 654-6266 • Fax: (609) 654-4309
E-mail: custserv@infotoday.com • Order online: www.infotoday.com