Digital Revolution:
Navigating the Process of Collection Digitization

Cristina Caravaca
University of Oregon
AAD Museum Studies

A capstone paper presented to the Arts & Administration Program in partial fulfillment of the requirements for the degree of Master of Arts in Arts Management.

June 2006

Digital Revolution: Navigating the Process of Collection Digitization
Submitted by: Cristina Caravaca
Date: June 2006

Approved by: ____

Dr. Janice W. Rutherford
University of Oregon
Arts and Administration Program
Museum Studies Concentration

Date: ____

ACKNOWLEDGEMENTS

I would like to thank my research advisor, Dr. Janice W. Rutherford, for her thoughtful critique of the drafts that preceded this final paper. My writing style has improved greatly under the gentle pressure of her editorial pen. I would also like to thank the University of Oregon Specials Collections Records Manager, Erin O'Meara. Her guidance during the literature review phase of this project was invaluable. Finally, I must extend my deepest gratitude to my family: Sandra, Alejandra, Gus, Zoe, and Otis. Each of them has sacrificed time, money and sleep so that I could pursue this degree. Without their unwavering support, I would never have reached this milestone.

TABLE OF CONTENTS

ABSTRACT

This navigation guide is intended for the museum practitioner who is about to embark on a collection digitization project, but has not yet established a protocol for performing the work required by such a project. After reading this paper, the project manager will have an understanding of the fundamental elements required by a digital conversion project. Areas addressed here are: benefits, the digitization chain, planning strategies, selecting material, content management issues, and the ethical concerns of a virtual museum. This guide assumes that the reader will have a basic level of technological proficiency.

INTRODUCTION

In a museum, collection digitization refers to the process of creating a computerized database to store provenance information and object images for the museum's holdings. Digitization is a multi-faceted process that demands expensive computer software and enormous amounts of staff time. This paper is a navigation guide that examines the basic steps required by the process of collection digitization. Its intent is to chart a course toward the ultimate goal of creating a functional database. This paper is not a "how-to" guide. I will not provide a step-by-step plan for the digitization project; rather I will provide the project manager with a conceptual framework for creating a path to the end product. There seems to be an endless array of digitization manuals available to potential project managers. However, all of these manuals are written for users who have a deep understanding of computer technology. This navigation guide is intended for small to mid-sized museums that are beginning to plan for their digital collection. The guide assumes that the intended audience has basic knowledge of computers, but also that they have very limited understanding of various proprietary software programs or the capabilities of peripherals such as scanners, cameras, or printers. After reading this navigation guide, the project manager will feel comfortable searching for a more complete instructional digitization manual.

This paper focuses mainly on image databases for two-dimensional and three-dimensional museum collections. It does not address the topic of electronic records management as it relates to personnel files, financial records, or any other non-image database topic.

BENEFITS OF DIGITIZATION

"*Digitization can allow wider access to fragile and rare materials. It can allow researchers to use materials from anywhere on the face of the planet, allowing physically disparate collections to exist in the same virtual space. It can even turn massive numbers of analog monographs, novels, magazines, and serials into searchable databases, allowing a completely new type of access to those texts."*

[Abby Smith in Kenny & Rieger's, *Moving Theory into Practice*, 2000, pg. 3]

The benefits of embarking on a digitization project are easy to envision. Whether the institution is a museum, historical society, or library, converting from analog to digital records allows collections managers to tame the paper tiger. Collections staff spend a large amount of their time caring for the papers that identify the material holdings of their institution. Constantly updating provenance files leads to bulging folders, and this leads to a cry for more physical space in order to accommodate the paper files. One obvious benefit of electronic records keeping is that no matter how large a provenance file gets, its electronic version will never take up any more space than the computer and server that host it. This aspect of digitization is so attractive that recent trends suggest that museums and libraries are moving towards an era when all object records will be created digitally, without ever having an analog form (Kenney & Chapman, 1996).

While an economy of space is the driving force for many digitization projects, wider access to collections information is another benefit. Collections managers can afford to be more generous with access to digitized record archives because potential users no longer need to intrude upon the physical space of the collection in order to search the archive. Safe access helps the museum welcome more patrons into previously inaccessible areas. "In the twenty-first century, with the development of digitization and the numerous online full-text databases, the possibility of researching from home once again exists" (Roff, 2005, p. 49). Records can be searched from remote locations, while their associated objects enjoy the security of museum storage. Often objects that are the most fragile are among the first candidates for digitization. While a digital photograph can never replace the original object, these surrogates can provide a view of something that might otherwise be restricted from viewing. "Cultural managers and staff attempt to preserve collection objects in superior condition for as long as possible. Every object finally changes. With a good conservation programme, managers extend the time that the object remains in good condition and extend its useful life for the purposes of exhibit, research, and education" (Fahy, 1998, p. 281).

Virtual museums

A digital collection that is posted on the World Wide Web can act as a global ambassador on behalf of its institution, allowing patrons to discover the holdings of an organization at any time of the day or night. Lesser-known organizations are now able to raise awareness of their collections in a way never before possible, and more prominent institutions can benefit from digitization by providing virtual access to the objects in their collection that are rarely ever on view.

Relating to the mission

When building a digital collection, one must always keep the purpose of the organization's original collection central to the project. It is neither practical nor useful to digitize every object in the collection. In every museum, there are objects that, while valuable, may no longer have direct relevance to the current mission of the organization. These objects might have been curated with another gift or donated in the founding days of the facility before a solid collections policy was in place. As an organization begins to decide which objects should receive priority treatment in a digital conversion project, the mission of the institution should shape the overall project. Guidelines about preservation, intended audience, educational focus, and vision for the future are embedded in the mission statement of the institution. These goals should infuse every decision made during the conversion. The image database has the potential to become the main access point into the collection; therefore, objects in the digital collection should support every facet of the organization's mission statement. Not only should they be relevant to the mission of the collection, but objects displayed electronically should also reflect the standards and overall content of the physical collection.

Serving the community

With so much information being loaded onto the World Wide Web, digital collection builders need to be diligent in avoiding redundancy in the objects they post on the Web (Murphy, 2003). Thoughtful consideration should be given to the selection of objects that will be publicly accessible. "Be certain that what you are offering is a discrete, cohesive, and hopefully unique collection of material that is in demand and has not already been duplicated on the Web" (Murphy, 2003, p. 155). Each object or document posted online should serve to broaden the field of knowledge, not just add to the sheer quantity of information already posted. The demand for online access to museums remains high among patrons, but an organization's portal will go unused if the information presented is confusing or poorly organized.

While the benefits of digitization are easy to grasp, the conversion process can be an obtuse and alienating experience. In attempting to discuss options with vendors or collaborators, the practitioner can become inundated with technological jargon. To make matters worse, many computer terms have multiple forms and meanings. In order to have meaningful conversations with technology specialists, the practitioner will need to acquire the basic vocabulary of digitization.

THE DIGITIZATION CHAIN

This section of the paper attempts to reduce the overwhelming abundance of terminology to a straightforward "digitization chain" that describes the five major system components and explains their functions in brief (Kenney & Chapman, 1996, p. 48). Recommendations made in this section have been chosen because of their widespread use. However, every digitization project is unique, and adjustments should be made by the practitioner as needs arise.

The five major components of the digitization chain are the central processing unit (CPU), monitor, scanner, storage area, and printer. Often referred to as the hardware, these are the machines that we can see and with which we interact when creating digital files. They are different from the software, which is actually a set of instructions that can be loaded onto the computer to perform specific functions. Software usually comes in the form of a compact disc that contains specific instructions for name-brand (proprietary) applications. Microsoft Windows XP and Adobe Photoshop are examples of proprietary software. The common word "computer" describes the system as a whole.

Central processing unit

The CPU acts like a brain for the computer. It carries out the functions requested by the user. When the user opens programs or searches files in the computer, the CPU performs the task of finding the requested item and sending it to the monitor. The CPU can also be called the microprocessor, or just the processor. Processors are measured in gigahertz (GHz). The term gigahertz describes how fast the processor can perform the function requested by the user. Intel Pentium, Celeron, or PowerPC are common brands of processors. An average speed for the processor is around 2.7 GHz (NINCH, 2003). Good computers have fast processors that can quickly locate the information that the user has requested.

In addition to an efficient processor, the CPU also requires adequate space for memory. In computer terminology, memory is described as Random Access Memory (RAM). RAM is the "primary storage area to write, store, and retrieve information and program instructions so that they can be used by the CPU" (Kenney & Chapman, 1996, p. 61). Adequate memory, measured in megabytes (MB), enables the CPU to handle scanning, editing, displaying, searching, and network transferring functions. Displaying and editing images requires a large amount of memory. As a rule of thumb, editing programs, such as Adobe Photoshop, have a RAM requirement of at least three times the image file size (NEDCC, 2000). For example, to edit a 7 MB sized digital photograph the CPU requires at least 21 MB of free RAM space in addition to the space allocated to storing the editing program software.

In order to plan for the acquisition of new hardware, the project manager must understand these basic requirements of computer operations. "The rapid pace of CPU development-and the heavy demands that new releases of graphics-based programs place upon processing power and RAM- more or less limit a computer's functional lifespan for high-end imaging applications to approximately three years" (Kenny & Chapman, 1996, p. 61).

Monitor

Unlike analog records, digital records are machine-dependent. Analog records, such as photographs, manuscripts, or microfilm can be viewed by the human eye. Digital records, such as those recorded on a compact disc (CD) or a digital video disc (DVD) cannot be understood in any tangible way. They rely on computer translation in order to be understood by humans. This dependency makes the monitor a key component in the digitization chain. Because digitization provides a surrogate for the original object, it is important to choose a monitor of the highest possible quality. Without a good monitor, it is impossible to create an accurate digital copy of the original object.

The quality of the monitor can be judged by its resolution and color reproduction. Resolution refers to the amount of information that can be represented on the screen. A monitor screen is made up of thousands of pixels. Pixels, also referred to as dots, are like ink spots on a newspaper page. Screens with more pixels per inch have sharper, clearer resolution (NEDCC, 2000). Conversely, screens that have fewer pixels available for resolution appear fuzzy. Aligned on an invisible grid, the pixels translate digital information into dots of colored light that the human eye interprets as text and image (NEDCC, 2000). Monitor resolution is measured in dpi (dots per inch) or ppi (pixels per inch). Technically speaking, dpi refers to printer resolution and ppi refers to monitor resolution, but these terms are often used interchangeably.

A monitor resolution of 72 dpi is sufficient for viewing, but not for printing (which will be discussed later). "A monitor's dpi is determined by the relationship between its screen size, which is fixed, and its pixel dimensions, which vary according to the setting chosen by the user" (Kenney & Chapman, 1996, p. 80). In order to determine the dpi for a given monitor screen, the horizontal pixel dimension is divided by the width of the screen in inches. For example, a 1024 pixel wide screen divided by 12 inches equals a resolution of 85.3 dpi. Since 72 dpi is about all the human eye can detect, 85 dpi will yield a crisp image. When choosing a monitor for the digitization project, one should not assume that every screen has a resolution of 72 dpi or better. Also, note that the manufacturer will measure the screen diagonally, but that the equation for screen resolution uses the horizontal width in inches to determine dpi (NINCH, 2000). A case in point: Toshiba advertises their Satellite laptop monitor as being 14 inches wide. This dimension is diagonal; horizontally it is only 12 inches wide.

Resolution, measured in pixels, tells us how well the monitor screen can emulate shapes and texture. Color, or bit-color, describes how many hues and shades the monitor can accurately reproduce. Monitor color is measured in bits. A bit is the smallest possible unit of computer data (Sharpened website, n.d.). “Color scientists estimate that the human eye can distinguish up to 10 million colors, but they also note that everyone perceives color differently” (Hunt in Kenney & Chapman, 1996, p. 86). Eight-bit color is adequate for black and white line drawings or pages of text, and sixteen-bit color accurately represents the many shades of grey. Twenty-four-bit color is more than enough color information to satisfy our eyes, but many computers now come with thirty-two-bit color, which is probably excessive. It is sufficient to use a monitor set to twenty-four-bit color for all viewing needs. Color richness is produced by the proximity of the pixels to one another. Each pixel can only be one color at a time, but when the pixels are placed side by side, they blend optically to register as a third color in our eye (NEDCC, 2000). Operating much like pointillist painting, a screen of pixels that alternate between red and yellow hue will appear orange to the viewer.

Today’s monitors are good at reproducing color, but the exact tone of any color is subjective. It is nearly impossible to match exactly the onscreen color of an image to the color of the original object. Even if a close match could be made on one monitor, every other monitor used to access the image would display a different shade of the same color. For example, blue tones in a digital image of a painting may appear green, turquoise, or even black depending on the color resolution of the monitor displaying the image. Should the practitioner decide to print the image seen on the monitor screen, another set of color considerations regarding printer accuracy must be addressed. "A Color Management System (CMS) should be employed by any institution wishing to accurately reproduce color from the source material (original object, artifact or film master) throughout the entire chain of digital hardware devices including monitors" and printers (NINCH, 2003, p. 113). However, if printing is a priority, it is best to solicit the services of a professional print shop.

Capture devices

In the language of digitization, to "capture," is to record the likeness of an object onto digital recording media. "Capture" can also refer to the way that contextual information, also called metadata, is gathered at the time of image capture. The most common devices used to capture images are scanners and digital cameras. When possible, it is best to capture the image directly from the original source, that is, photograph or scan the original object or document. "With any intermediary there will always be some information loss, so digitizing from the original is advisable when the highest level of quality, per item, is needed" (NINCH, 2003, p. 104). In the case of exceedingly brittle objects, it may be preferable to digitize from an intermediary medium (i.e. negative, slide, or photograph) rather than expose a fragile object to the rigors of scanning.

Digital cameras

Digital cameras work on the same principle as a traditional camera, but rather than analog film, the camera saves images to an internal digital recording medium. Often referred to as cards or sticks, these recording devices are usually no bigger than 2 x 2 inches and some can store up to 300MB of information. They can mimic the proportions and resolution of analog film convincingly, and, in addition, the recording device can be erased and re-used several times.

Common brands of recording devices are FlashCard, Memory Stick, and MultiMedia Card. The resolution of a digital camera is measured in megapixels. A megapixel is a unit of measurement that equals one-million pixels (Sharpened, n.d.). The more megapixels, the more information about the image or object the camera can record. A digital camera of five megapixels can easily produce an image that is larger than the standard 1024 x 726-pixel computer screen.

Digital cameras are very useful when the original object is an awkward shape, such as statuary, or when the original is too large or too fragile to be scanned. The main drawback in using a digital camera to capture the image is that it is difficult to control lighting and background clutter in the final photograph.

Scanners

Scanning always produces digital images of better quality than digital photography. "Scanners operate in much the same fashion as traditional cameras, and share many of the same features, beginning with the optics. Each device utilizes a lens to focus the incoming light, an aperture to regulate the amount of light, and a shutter which determines the length of time the light is allowed to shine through the aperture. Each device also contains a recording medium on which the incoming light is focused and the image is formed" (Kenney & Chapman, 1996, p. 51). There are several types of scanners on the market. The most common are flatbed scanners, drum scanners, and slide scanners.

Flatbed scanners are the most commonly used scanner. They are affordable and can accommodate flat objects up to 11" x 17". They resemble photocopiers, but are smaller and flatter. "Chief differences between high-end and low-end flatbed scanners are speed of throughput and software enhancement capabilities" (Kenney & Chapman, 1996, p. 53). Flatbed scanners can have a single source light, a light that shines through the scanning plate from the bottom, or dual source light shining from both the top and bottom onto the original. Dual source flatbed scanners work well with transparent media, such as negatives or slides.

Drum scanners were originally designed for the graphic arts field and are very rarely used in anything but a professional print shop. As the name implies, these scanners have an image-capturing drum to which the original object must be attached before it can be captured. The drum scan process can be detrimental to brittle objects. Although they have the capacity to capture extremely precise images, drum scanners "are very expensive and impractical for large projects" (NINCH, 2003, p. 109).

Slide scanners are specialized devices that are shaped specifically to accept slide media. They are still widely used; however flatbed scanners with dual source-light are beginning to outperform them.

When choosing a scanner for the digitization project it is important to select the scanner with the largest scanning plate (no less than 11" x 17") and the highest resolution rate one can afford. The scanner will also come with a software package. "Lower-end scanners are often sold with software of limited functionality, which might require using an additional software package" in order to achieve good image quality (Kenney & Chapman, 1996, p. 58). If resolution and scanning plate size are equal, choose the scanner that has the software package that includes the best editing capabilities.

Storage systems

Best practices for scanning recommend that an image be captured at the highest resolution possible. "Under most circumstances, it would be a waste of time to scan the new materials at an arbitrarily low resolution" (NINCH, 2003, p. 108). The museum should strive to create a large, digital master image for long-term preservation. The digital master should represent the original as accurately as possible and should be "raw," which is to say, not manipulated through editing software. This ensures that the file format of the master will be applicable to a wide range of future uses. Once a master has been created, subsequent smaller, web-formatted, or printable images can be made from the master without handling the original object a second time (Kenney & Chapman, 1996).

Types of storage

There are two broad categories of file storage: online & offline. Networked computers are those that are connected to one another through a server. The server is another type of computer that is dedicated specifically to the task of handling and storing files for the computers in its network (Quinn, 2006). Online storage allows each computer to save data to the network server. This type of storage protects the recorded data from the possibility that an individual computer will become unstable and lose information. It also allows the user to access information saved to the server from any computer in the network. Online storage should be managed by an Information Technology (IT) specialist.

Offline storage is storage on any kind of external media that is not backed up to an online server. A CD of digital images sitting on a shelf is an example of offline storage. This type of backup allows files to be stored without allocating any of the computer's memory for storage space. Data stored offline can also be transported to computers that do not have network access (NEDCC, 2000). Most systems use a combination of online and offline storage to perform their functions.

The final component in the digitization chain is the printer. Regardless of how much care has been taken to create high quality images, or to match the color of the original to the color on the screen through CMS practices, everything can be wasted if the printer is substandard.

Consider the many layers between original, digital copy and print-out. First, an image of the original must be captured, either by camera or scanner. Second, the captured image is transported through wires, saved in the CPU, and then reassembled on the monitor screen. Along the way, there are many opportunities to lose or misinterpret information about the color and texture of the original, analog object. If the image must travel to the printer, it will encounter another set of choices regarding color and texture representation. Even the type of paper used will effect the resulting reproduction of the original. Despite the utmost care, “monitors and printers may distort image quality” (Kenney & Chapman, 1996, p. 76).

Printer

Non-commercial printers come in two basic varieties: inkjet and laser. Inkjet printers are the most affordable and common of the two. They spray fine particles of ink onto the paper. Low-end inkjet printers can produce only low-resolution images that have a grainy appearance. High-end inkjet printers do a better job, especially when used with good quality paper. Laser printers offer higher resolution and better consistency of color. They project a very fine powder toner onto the paper, which is than affixed by heat and electromagnetic charge (NEDCC, 2000). Laser printers are also more expensive to purchase and operate. Both styles of printers measure their resolution in dpi, which can range from 300 to 3000 dpi. The higher the dpi the smoother and more detailed the printed image will be. Image print-outs should be used as references only. For distribution-quality printed images it is best to work with a professional printing vendor.

Each of the five components of the digitization chain (CPU, monitor, scanner, storage, and printer) is equally important in the creation a digitization workstation that is both effective and efficient. Understanding and obtaining good quality hardware is an important element in the digitization project. However, in order to manage a successful digitization project, it is important to forecast a plan for the conversion process.

PLANNING FOR SUCCESS

"Setting goals represents the thinking or brainstorming first phase of a project, and a good manager knows when to make the transition to planning, the second phase" (NEDCC, 2000 p. 36). The previous sections of this paper have addressed the benefits of digitization, as well as some of the technological requirements of the project. Now it is the duty of the project manager to create a roadmap that will guide the organization through the conversion process.

In-house vs. outsourced production

Before an organization can establish any kind of action plan for digitization, it must first decide who will do the actual work of digitization. The process of converting object records from analog to digital form is composed of many steps. The objects must be selected, removed from their resting place, captured digitally, described, and documented. The record is then created, formatted, and saved to a back-up file (Hunter, 2000). Finally, the original objects must be returned to their original location. All this manipulation requires human work hours, so clear decisions must be made as to just who will provide the labor.

Many museums choose to do the work "in-house," meaning they utilize their own space and personnel to meet the needs of the digitization project. The benefits of in-house production are many. An in-house design means that fragile objects need not leave the security of their facility. It can provide an opportunity for collections staff to increase their knowledge of the chosen objects while simultaneously becoming acquainted with the newest developments on the technological front. If the project is eventually outsourced, in-house design can also be a first step toward an effective vendor-client relationship. Managing its own digitization project means that the institution has full and complete control over every aspect of the digitization process. The organization has the freedom to progress at the rate it chooses and even to switch course in the middle of the project. However, it can also drain attention away from other areas of collections management (Smith, 1999).

Digitization projects demand space for computers, scanners, and photography staging areas; they also require personnel with specialized training who can operate the equipment. Outsourcing is a viable choice for an institution that cannot provide the staff or space necessary for an in-house production. To outsource a digitization project is to hire a professional digitization vendor to perform the functions necessary for project completion. The vendor and the institution work together to establish end goals, time frame, and budget constraints for the project. The vendor will usually take the agreed upon objects to a professional studio, photograph them and send the work back to the institution for review (NEDCC, 2000). Once the images are approved, the objects are returned to the institution. Although it may seem attractive to simply have the finished product appear on your doorstep, the dangers of transporting objects to a vendor's studio and back again are very real and should be taken into consideration.

Economic advantages are potential benefits of working with digitization vendors. Considering the rapid rate at which computer systems become obsolete, utilizing the equipment and "economies of scale" of a vendor might save the organization some money (NEDCC, 2000, p.152). However, even if the digitization project is outsourced, the organization will still need to provide adequate equipment in order to access and maintain the digital database constructed by vendor.

Selecting material

When faced with the decision of just what to select for digitization, many institutions feel that the best solution is to digitize everything. This is both impractical and potentially destructive to the efficiency of the project. To select everything for digitization would elevate cost and lengthen production time, making the project impractical. However, it is a difficult challenge to choose the parts of the collection to be digitized: "[S]election criteria for digital conversion are so varied and overlapping that it is difficult to organize them and assess them coherently" (Kenney & Rieger, 2000, p.11).

User demand

One of the most popular methods for determining what to digitize is to choose objects by user demand. Because the digitization project as a whole is a user-demand driven endeavor, it makes sense to refer to the potential user as a guide. Under this criterion, digitization projects should begin with objects that are already in high demand at the institution. For example, if a museum has a well-known piece of Greek pottery in its collection, then the famous pot should be among the first digitized objects. Next, items relating to the pot can be added, in order of user demand, to create a more complete representation of the museum's collection.

Another user-driven selection theory is to try to anticipate the next trend in user demand. On this subject, Kenney & Rieger (2000) write, "Intellectual content may be the overriding causal factor when deciding to digitize a collection, especially if current research trends and demands conspire to make it a timely topic of inquiry" (p. 15). In this model, organizations try to anticipate what the user will be researching in the near future. Heightened interest is often spurred by the anniversary of a historical event or the birthday of a historic figure. While this may be somewhat subjective, it is also a great opportunity for the organization to raise awareness of its ongoing digitization project. Having selected the material to be digitized, some decisions must now be made as to what kind of digital image is desired. "We advocate an approach to digital conversion that is designed to ensure high quality and functionality while minimizing the costs" (Kenny & Chapman, 1996, p. iii).

Finally, when choosing the primary objects for a brand-new digitization project, start with objects that seem easy to capture. Choose an object that is in a state of conservational stability rather than delicate embrittled objects. Also, choose objects that can be scanned easily rather than beginning with odd shaped or oversized objects. As the project progresses and the skill of personnel increases, things that once seemed difficult will become more manageable. By beginning with simpler objects, you allow time for adaptation to new protocols and new equipment.

Timeframe for the project

One of the first things any project manager worries about is the timeframe of the project. Although important, time spent planning, selecting materials, and making decisions about quality is invisible to those outside the project. "There may also be pressure to hurry this [planning] step, from a desire to show visible progress or in response to institutional pressure" (NINCH, 2003, p. 9). There is a temptation to rush through the preliminary stages in an effort to generate a viable work product. However, shortcuts should be avoided. A well-crafted plan will keep the project on task and save production time in the long-run. The project plan should include some flexible time in order to allow for refinements to the digitization plan. "Expect the first few months to be a shakedown period" (NEDCC, 2000, p. 162). Open communication often relieves some of the pressure a manager may feel about timelines and the rate of work production. The manager should create a map that denotes the stages of the digitization project. The map can then be used to share information with other members of the institution.

Phases and batches

Because the digitization project can be complex, it is often best to reduce the action plan to phases and batches of activity. Phases of the project refer to the larger activities, such as selecting objects, planning a timeline and establishing a digital workstation. Batches are the smaller steps that make up various phases. A project manager working in the selection phase for an art collection may reduce the selection phase into smaller batches by art medium. For example, the manager may select paintings for digitization first, and then drawings, and then sculptures and so on until the selection phase is complete. Phases and batches are not technology oriented terms; rather they are ways of framing the actions that will take place during the project (Hunter, 2000). What is most important is that the plan follow a natural flow so that no object is disturbed unnecessarily, or removed from the storage space for extended periods.

Threats to the collection

Careful consideration of the possible threats to the areas of bibliography, inventory, preservation, and physical security of the objects will go far in aiding the organization in creating a successful digitization project. This navigation guide recommends a risk-management approach to security in order to address these major threats to the digitization project. This is to say that, in order to plan for success, one must imagine the possibility of failure (Giesecke, 1999). "To understand where [a digitization project's] risks lie, management must be clear about what risks might threaten the mission of the institution" (Price & Smith, 2000, p. 4). It is useful to envision the physical forms that failure may take, and where the loss might occur. "The four salient types of risk include bibliographic, inventory, preservation, and physical security" (Price

& Smith, 2000, p. 8). A case-specific protocol should be established to prevent losses in these four main areas.

Bibliographic threats are those that would cause a loss of information that describes the object and its history. Bibliographic material includes information such as the name of the artist, important dates, significant locations, and relating materials. A failure at the bibliographic level may mean that some piece of the information is incorrectly entered into the digital database or that the information is lost altogether (Price & Smith, 2000).

An inventory threat is one that threatens the tracking capabilities of a database system. Inventory refers not only to the holdings of an institution, but also to the location of the objects listed. The inventory information about the object must be carefully controlled during the transportation of the object from its storage or exhibition location to the digitization studio, and back to its original location. This process may take several days, so a log of location and activity associated with the object can be helpful in maintaining an accurate inventory. Inventory control can also fail at the data entry level. Inventory lists can record loan histories, conservation measures, exhibition status, and a host of other activities. Precautions must be taken to ensure that correct and accurate information is provided regarding the temporary and permanent locations of the object being digitized. A digital database that fails to provide accurate and timely location information is a threat to the overall quality of the digitization project.

A threat to the physical security of an object is any overt, malicious attempt to damage or steal the object. To guard against this threat, the museum must consider the history of any worker who is involved in the digitization project. Digitization is a labor intense undertaking, and many organizations call on volunteers to supply some of the work hours necessary for the project. Using volunteer labor is a viable option for in-house digitization projects, but care should be exercised when choosing volunteers to perform specialized functions. Typically, volunteer staff members have not received the thorough background-check of a paid employee, and therefore only those that have been with the organization the longest should be put in charge of handling objects. Issues of privacy and security also surround the data entry of provenance information connected with the object. Workers should be briefed on the importance of keeping certain aspects of object location and donor information private. The actual building or part of the building, where the digitization is taking place might also pose threats to physical security. The project manager needs to establish a protocol for temporary staging areas, temporary storage areas, and transportation routes between the permanent location and the digitization studio. It is also wise to create a list of people who have access to the objects, as well as those who can travel beyond the boundaries of locked doors.

Preservation threats endanger the institution's ability to care for the object in perpetuity. During the digitization project, the threat is to the original, analog object; after digitization is complete the preservation threat is to the sustainability of the database. Database sustainability is discussed in the content management portion of this paper.

Preservation of the physical object is a central concern of any collections division. Objects in storage and on view must be guarded against the effects of light, temperature, humidity, pollutants (air quality), insects, mold, mildew, and human error (Bachmann, 1992). These elements are difficult to control under normal circumstances, and the potential for damage by any one of these threats becomes even greater during the digitization process. In order to digitize an object, the object must be removed from its safe environment, transported to another place, handled, and exposed to the strong light necessary for photography or scanning. None of this is avoidable if a successful image capture is to occur, but the constant threat to appropriate preservation should never be far from the project manager's mind. The key to minimizing preservation threats to an object undergoing digitization is careful forethought. A work plan can help reduce accidents by helping the workers visualize potential accidents that might cause damage to an object.

Once the timeframe for the project is established, and the objects to be digitized have been selected, the project manager must consider the way object information will be stored and displayed in the database. Often this step of the planning process is rushed, resulting in a database that fails to live up to its full potential. Abby Smith (2001) writes, "Much less has been written about how to plan for the access to and preservation of digitally reformatted collections over time than about how to select materials for digitization. This is partly because we know nothing certain about maintaining digital assets over the long haul" (p. 24). The next section of this navigation guide addresses protocol for establishing a sustainable digital database.

CONTENT MANAGEMENT ISSUES

Content management, as the name implies, is the practice of organizing the content found within a digital collection database. A typical digital collection will consist of two parts: an image bank and a database. The image bank stores only the image files for the collection. The database stores the contextual information or metadata for the images. Digital collection software allows the two parts to link together and create the unified appearance of images and accompanying text. Even a small digitization project can quickly generate hundreds of images and text files. "Without a browsable or searchable catalogue, end users will struggle" in their effort to locate objects within the collection (Lee, 2001, p. 103). Effective content management protocol will carefully consider file name structure, metadata standards, and migration practices.

Good management will also plan for the unavoidable obsolescence of file formats, proprietary software applications, and hardware upgrades.

Files

A file is any set of digital information that is saved in electronic format, such as word documents, tables, or images. Files are measured in bytes (B) and increase by increments of 1,024 units from byte, the smallest, to yottabytes, the largest (Sharpened, n.d.). The most commonly seen file sizes, in order of size, are: Byte (B), Kilobyte (KB), Megabyte (MB), and Gigabyte (GB) (NINCH, 2003).

The file format is the language that the file speaks when it communicates with the CPU. Some file formats capture information without loss of information. They are called lossless files. "The most widely adopted format for storing preservation quality digital masters is uncompressed [tagged image file format] TIFF" (NINCH, 2003, p. 105). TIFF files are lossless and are recommended for creating master files. Other file types compress the data in a way that results in lost information and lower resolution. These files are known as lossy files. The advantage of a lossy file is that it takes less memory space than a lossless file. Joint Photographic Experts Group (JPEG) is a widely used type of lossy file format that yields good quality photos for onscreen viewing, web posting, or small print-outs. When generating digital images, it is best to capture the archival master image in TIFF form, and then use JPEG as the file format for all derivatives from the master. A JPEG image, set to the proper resolution, can provide ample information for either printing or on-screen display.

File naming conventions

The need to name files arises at every stage of the digitization project. Therefore, a well-designed content management plan should provide a structured file naming system. Each object should be assigned a file name that is unique, persistent, and meaningful to the associated object. File names should also indicate that the file is part of a larger identification system (Buck & Gilmore, 1998). There should be “a catalog structure of varying levels and a structure for describing each individual object in the collection” (Lee, 2001, p. 104). In addition to being unique, the file name should also be persistent. That is to say that the name should remain the same no matter where the file appears (NISO, 2004). This is especially relevant when files are being saved into different formats. For example, the average digital camera assigns a sequential file name, such as DSCN0271, to every photograph it records. Here, DSCN is the name assigned to the JPEG file by the camera, and 0271 means that this camera has taken 271 JPEG pictures. The file name is unique to this image, but it is not meaningful to the end-user because the file name has no relationship to the collection. When this file is saved to the image bank, it should be assigned a more meaningful name. Let us say that DSCN0271 is one image from a sculpture collection by Rodin. The file name should use “descriptors that already exist” in order to relate the digital copy to other analog materials about the work (Lee, 2001, p. 107). By using existing descriptors, a better name for DSCN0271 might be Rodin_1912_torso. This file name informs us of the artist, date of work, and subject matter even without opening the image file.

Once assigned, the file name must be persistent. That is, it must remain with the image file in spite of changes made to the file (NISO, 2004) . If the image is manipulated with image editing software, we can indicate changes with a simple “_v1” at the end of the established file name, i.e. Rodin_1912_torso_v1 (NINCH, 2003). If this image is saved into different file formats, we can keep the file name the same and let only the suffix change such as, Rodin_1912_torso.jpg or Rodin_1912_torso.tiff. No matter where this file is used, the file name provides a path back to its collection.

When choosing a file naming convention, it is important to consider not only what the file name will convey intuitively, but also how files will be displayed (NISO, 2004). Computers operate under an alphanumeric discipline. That is to say that they organize information by letter and number. In order for groups of files to remain together when displayed as a list, they must share a file naming structure. Regardless of when the file is saved into the image bank, the computer is automatically set to re-sort the list by name, and always show groupings of common items. For example, all the Rodin files in our image bank will appear together first by Rodin, then by the date, then alphabetically by the subject matter word. Examples might be:

Rodin_1910_standingman.jpg
Rodin_1912_torso.jpg
Rodin_1912_untitled.jpg

In the beginning, this may seem somewhat tedious and labor intensive, but as the digitization project expands to include more media and broader time periods, the need for concise file naming and a structured folder system increases exponentially.

Digital master

The digital master should represent the original as accurately as possible, preferably in TIFF file format. The master should be raw, which is to say it should not be manipulated through editing software. This ensures that the file format of the master will be applicable to a wide range of future uses. Once a master has been created, subsequent smaller, web-format or printable images can be made from the master without handling the original object a second time. This process is known as scaling (Kenney & Chapman, 1996).

A "preservation quality digital master" should be created in a non-proprietary format (Smith, 2001, p. 7). Even when an image is edited with a proprietary software application such as Photoshop (.psd) or Adobe Illustrator (.ai), the image should still be saved in a non-proprietary format to facilitate future data migration. A TIFF file format with a minimum of 300 dpi resolution is the recommended format for archival images. Archival images should be saved separately, and not linked to the database (Lee, 2001). While TIFF format is excellent for preserving rich layers of information about an image file, it is too large to be used as an onscreen display. Instead, a lossy JPEG file format set for high quality resolution is a better choice. High quality JPEG images "support twenty-four-bit colour depth, but allow for compression (thus making the file size smaller)" (Lee, 2001, p. 46). JPEGs provide ample information for the human eye, without carrying extra digital information that will slow the downloading time.

Image bank

The image bank is the area of the computer that is designated to hold digital images produced during the capture phase of the project. The image bank is nothing more than a folder on the computer's hard drive that has been established for the express purpose of storing images related to the digitization project. For example, the Microsoft Windows computer application provides a default, image bank folder titled, My Pictures, with its software package. It is good content management practice to keep the image bank for the digital collection separate from other image banks stored on the same computer or network (Smith, 1999). Having separate image banks for different image collections facilitates access. If the computer storing the image bank is linked to a network server, then the image bank folder should be saved onto the network. This protects the image bank folder from being destroyed if the computer becomes unstable. It also makes the images from the digital collection available from multiple computer terminals.

The database

The only function of the image bank is to hold images for the digital collection. The database is responsible for storing contextual information or metadata about each specific image.

These two content areas working together create what we think of as the digital collection. Typically, a museum will invest in a proprietary software application to act as the database. PastPerfect, Microsoft Access, and The Museum System (TMS) are examples of commercially available image database software applications that have the capability of storing information and linking it to digital images. The advantage of a specialized database is that it stores metadata information in a standardized format for every object.

Metadata

Metadata is the contextual information about the object being digitized. It can incorporate any information the project manager deems relevant to the object. There is a careful balance to be struck between recording thorough and complete metadata, and remaining within the boundaries of the timeline of the digitization project. Because the information that comprises metadata is subject to the desires of the organization, efforts have been made to standardize the information recorded by metadata records. The Dublin Core Metadata Initiative (DCMI) was established in response to this need. DCMI is an "organization dedicated to promoting the widespread adoption of interoperable metadata standards and developing specialized metadata vocabularies for describing resources that enable more intelligent information discovery systems" (Dublin Core n.d.). In order to be Dublin Core compliant, a digital collection must provide, at a minimum, the following fields of information for each object in its collection: title, creator, subject, description, publisher, contributor, date, type, format, identifier, source, language, relation, coverage, rights (Dublin Core n.d.). The presence of these fifteen core elements enhances the navigability and interoperability of the digital collection.

Case study: PastPerfect

PastPerfect Museum Software (created by Pastime Software Company) is one of the most popular digital software applications on the market today. It can hold thousand of records, offers multiple metadata formats, and is supported by a robust vendor that offers frequent upgrade packages.

To create a file with PastPerfect, the project manager must first create a digital image. Let us say that we are entering a record for an archival manuscript. First, we need to create a digital copy of the manuscript. We can do this by scanning or photographing the work at the best quality possible while still considering the safety of the object itself. Once the image has been created, it is saved to the image bank. Now we are ready to open PastPerfect and enter all known metadata for the manuscript. In PastPerfect, we can choose from several different data entry forms such as, art, archaeology, geography, or history. Each form offers a distinct set of questions about the object in question. After choosing the appropriate form, we simply fill in the blanks provided by PastPerfect. When all metadata has been entered, we can choose to link an image to the record. To do this we must tell PastPerfect where to find the image. The program will ask where the image is located. Using drop-down menus, we navigate back to the image bank where the digital image of the manuscript is saved and select it. Now PastPerfect knows which image is related to the file and it creates a permanent link. As long as the image bank is available, the program will always attach the digital image to this record in the database.

Migration

Every digital collection is vulnerable to obsolescence. To combat this dilemma, digital collections managers employ a strategy known as data migration. Migration "involves transferring files from one piece of hardware/software/file format to another" (Lee, 2001, p.

146). This step is recommended when newer versions of software or hardware become available and begin to replace the older versions. Data might also be migrated to avoid harmful "bugs" in the software. Data must be frequently migrated if they are to avoid becoming obsolete on the shelf. Copying word document files from Microsoft 97 into Microsoft 2000 is one example of migration. Another type of migration is moving data from one medium to another, such as from CD-R to DVD (NINCH, 2003). If migration fails to occur for any given file, the language necessary to read it will eventually be lost and the information contained within the file will become inaccessible. The process of "migration is labor intensive, requiring thorough planning and analysis, and it does not scale well. Each cycle may be a new experience, making long-term preservation planning difficult" (Kenney & Rieger, 2000, p. 147-148).

Content management dictates how effectively the newly created digital database will function. A well structured database can aid staff in performing their curatorial duties and also pave the way to an easy transition onto the World Wide Web.

ETHICAL CONCERNS FOR THE VIRTUAL MUSEUM

Patrons expect that a museum will have an internet presence. They use it to get the facts about location, hours, and ticket prices. They also expect to be able to see samples of both permanent and changing exhibitions, as well as upcoming events. Web users demand that the information they are seeking come to them in an easily digestible form (Skramstad, 1999). Users walk a line between wanting as much information as possible about every object, and simultaneously desiring small and manageable quantities of information for the object they are researching. One way to achieve this balance is to "avoid the thousand-pages-of-links syndrome" and instead provide a topic area menu that will allow a user to choose links by area of interest (Hall, 2002, p. 3). The topic area model guides the user on an intelligent query. It is preferable to the "thousand-links" model because it requires less knowledge on the part of the user. For example, imagine navigating the homepage of a museum. You could be presented with the name of every artist in the museum's collection, or you could be offered a search that is based on the library catalog paradigm (Hall, 2002). Using a library catalog paradigm, you would choose general areas (e.g. by genre or date) and then allow the computer to perform the search according to your restrictions. The latter has more potential to offer interesting cross-referential material because the user has chosen the main topic area. A laundry list of every artist in the house does not offer an informed browse and is more likely to turn the user away from the institution's virtual doorstep. The digital collection builder should consider the needs of the potential user when creating the design of the database (Hall, 2002).

Database case study: MoMA

The Museum of Modern Art (MoMA) in New York offers an excellent example of an effective digitized collection's public portal. This image database is easy to navigate and allows the patron to search in several different ways to find information. MoMA's homepage (www.moma.org) offers a plethora of information about upcoming shows and events, related art venues, physical location, and contact numbers. A navigation bar at the left of the page offers a menu of links that lead to other active areas of the website. By following a link called "collection," the user arrives at the collection's database homepage. In addition to some general information about the permanent collection, a user is given three main search options: search, search and browse the online collection, or choose by genre. Following the link of choice, the user is presented with a more refined set of options. The link for the genre "drawing" leads the user to contextual information about drawing at MoMA, along with the further refined choices of: search, exhibitions and projects, drawing study center, highlights, or all works. Clicking on "all works" brings up page 1 of 105 thumbnail digital images of drawings that are in the permanent collection of MoMA. Because the user has followed informed links, this hundred-long list is likely to be a valuable search result. Choosing any of the thumbnails brings a larger image of the artwork along with metadata about the artist, locality, period, and so forth. Many of the entries are sensitive and link to other sources of information, which allows the user to cross-reference by serendipity. The MoMA website is an excellent model of a navigable sight that presents a wealth of information in small, digestible portions.

Creating a digital avatar for your museum is a powerful tool, but taking your place on the World Wide Web means that you are vulnerable to criticism by people who are far removed from your institution. The areas that tend to prove most troublesome for virtual museums to manage are decency standards and copyright law.

Decency standards

Much of the subject matter of fine art challenges the taboos established for our society. In many ways, it is the role of the artist to cause us to reflect on our own standards. Posting images that refer to controversial material can cause a negative reaction from patrons who may be uninformed about the nature or intent of the work. The 1973 court case of Miller v. California stated that material should be judged "obscene according to the standards of the local community" (Lee, 2001, p. 149). In this era of the global network the "local" community is far reaching, and therefore decency standards are hard to agree upon. The virtual museum must decide to engage in either self-censorship or risk being labeled as indecent (Quinn, 2006). The museum must also protect the images it posts on the Web from piracy by those who have no permission to use them.

Copyright issues

The content of a museum's website will contain copyrighted material in the form of intellectual property. "Essays, poems, computer programs, scientific studies, songs, paintings, and photographs are all examples of intellectual property" (Kaye & Medoff, 1999, p. 340). A website is vulnerable to researchers who might use and publish the findings of other scholars without giving proper citation or to those who would download images for redistribution without obtaining written consent. These unlawful acts hinder the ability of an organization to share knowledge about cultural artifacts. The museum should also be aware that there are times when publicly posted material may be used without the consent of the copyright holder.

While the Copyright Act of 1976 "gives copyright protection to original works of authorship", Section 107 of the act allows "fair use" of the same material for educational purposes (Kaye & Medoff, 1999, p. 341). This is to say that intellectual property is not extended copyright protection in the event that someone is using it to enhance his or her knowledge of a topic. Section 107 does not give the researcher the ability to publish excerpts or ideas without citing the original source (Quinn, 2006). Because the mission of most museums is to educate the public, it is difficult to avoid the risk of patrons violating copyright law when posting scholarship.

Conclusion

Before embarking on a digitization project, it is advisable that the project manager understand the basic tenets of the digitization procedure. He or she must be able to estimate the route the project will follow. The manager should also have at least rudimentary knowledge of the technology required by the project and be able to anticipate some of the pitfalls that may be encountered on the road to digitization. Finally, having successfully created a digital collection, the manager should be able decide whether launching the collection onto the World Wide Web is beneficial to the mission of the museum.

Computers are not magical machines that effortlessly produce good quality photographs and neatly standardized metadata. A functional digitized collection is the result of countless hours of intense labor by the museum staff. "Digital resources … are seen as objects that can be easily recreated — one only need scan the picture again. Fortunately this lackadaisical attitude appears to be diminishing in most organizations and in its place is the growing belief that digital resources are, at the very least, as valuable as the time, effort, and finance that has gone into their creation" (NINCH, 2003, p. 149). The project manager must create a methodical and detailed plan of action in order to turn the dream of a digitized collection into a reality. This navigation guide provides only the basic conceptual framework on which a digital project can be based. With this framework in mind, the project manager should feel confident in delving into more detailed digitization instruction manuals.

REFERENCES

Bachmann, K. (ed.). (1992). *Conservation Concerns: A Guide for Collectors and Curators.* Washington, DC: Smithsonian Institution Press and Cooper Hewitt.

Buck, R. & Gilmore, J. eds. (1998). The new museum registration methods. Washington, DC. American Association of Museums.

Dublin Core Metadata Initiative website (n.d.). Retrieved February 23, 2006 from http:// dublincore.org/.

Fahy, A. (1998). Collections management. London & New York, Routledge.

Giesecke, J. (1999). Scenario planning and collection development. *Journal of Library Administration*, (28)1, 81-92.

Hall, P. (2002). How we stay ahead of our Web users. *Computers in Libraries*, (22)1.

Hunter, G. (2000). Preserving digital information: A how-to-do-it manual. New York: Neal-Schuman Publishers.

Kaye, B. & Medoff, N. (1999). The world wide web: A mass communication perspective. Mountain View, CA: Mayfield Publishing Company.

Kenney, A., & Chapman, S. (1996). Digital imaging for libraries and archives. Ithaca, NY, Cornell University Library.

Kenney, A., & Rieger, O. (2000). *Moving theory into practice: Digital imaging for libraries and archives*. Mountain View, CA: Research Libraries Group.

Lee, S. (2001). Digital imaging: A practical handbook. New York: NY, Neal-Schuman Publishers, Inc.

MoMA website (n.d.). Retrieved January 9, 2006 from http:// www.moma.org/.

Murphy, J. (2003). Link it or lump it: basic access strategies for digital art representation. *Journal of Library Administration*, (39)2/3, 139-160.

National Information Standard Organization (NISO) Framework Advisory Group. (2004). A framework of guidance for building good digital collections. 2nd edition, *NISO Press*, Bethesda: Maryland.

National Initiative for a Networked Cultural Heritage (NINCH). (2003). The NINCH guide to good practice in the digital representation and management of cultural heritage materials. Retrieved 9 January 2006, from http://www.nyu.edu/its/humanities/ninchguide/.

Northeast Document Conservation Center (NEDCC). (2000). Handbook for digital projects: A management tool for preservation and access. Adover, MA, NEDCC publication.

Price, L. & Smith, A. (2000). Managing cultural assets from a business perspective. C*ouncil on Library and Information Resources*. Washington, DC.

Quinn, M. (2006). Ethics for the information age. Boston: MA, Pearson Education, Inc.

Roff, S. (2005). The return of the armchair scholar. *Journal of Scholarly Publishing*, (1), 49- 57.

Sharpened Website (n.d.), *Glossary*. Retrieved March 15, 2006, from http://www.sharpened.net/glossary/.

Skramstad, H. (1999) An agenda for American museums in the twenty-first century. *Journal of American Academy of Arts & Sciences* (128)(3), 11-30.

Smith, A. (1999). Why digitize? C*ouncil on Library and Information Resources*, Retrieved 9 January 2006, from http://www.clir.org.

Smith, A. (2001) Strategies for building digitized collections. C*ouncil on Library and Information Resources*. Washington, DC

Wilhelm (2004). Digital nation: Toward an inclusive information society. Cambridge, Massachusets: MIT Press.