
Securing Digital Image Assets in Museums and Libraries: A Risk Management Approach

TERESA GROSE BEAMSLEY

ABSTRACT

THERE IS AN OBVIOUS NEED FOR ONGOING RESEARCH, evaluation, and planning if museums and archives are committed to protecting their digital image assets. A number of potential threats to the integrity of digital image information can be identified when standard practices in museums and archives are examined. Changes in the integrity of digital image information can be caused by the manner in which the source data are acquired and recorded and by modifications made to the image data file. Alterations made to contextual data can limit valid interpretation of the associated surrogate image. The destruction of the mechanisms that link contextual data to the appropriate digital image has the same effect as deleting contextual information. Loss of control over digital assets can be the result of failure or inability to establish and publicize copyright. Even if copyright is established and enforceable, failure to enforce rights has the same effect as having no rights at all. Finally, failure to detect corruption of digital information means that invalid, partial, or inappropriate information will be spread under the guise of authentic reliable information.

Some institutions are already proactively applying security measures to digital image collections. Some of these security measures can have a negative impact on the integrity of the files that they are designed to protect. Systematic consideration of risk factors can inform the creation of procedures and application of security that works to guarantee the reliability and accuracy of digital image assets.

DIGITAL IMAGE INFORMATION AS AN INSTITUTIONAL ASSET

In their earliest manifestations, museums and archives were essentially collections of primary source materials. The collectors determined the criteria by which artifacts or manuscripts were chosen for preservation. The criteria were based at least in part on the value of the information that was embodied in the content of the materials or implied in the existence of the objects, a value that was established by the needs and interests of wealthy collectors.

Public exposure to museum and archival collections began in earnest at the turn of the nineteenth century. The infrequent opening of personal holdings to scrutiny became more commonplace as the general population came to recognize the existence of these collections and to demand access. In some cases, the profit motives of collection holders played a significant role in the growing accessibility of collections. The public saw value in the experience of gaining physical access to rare and unusual materials. The collectors saw value in offering access (sometimes for profit) to a new market. Selection of materials and the determination of their intrinsic information value were still determined by gentleman collectors. Increasingly, scholars used the information in their studies and augmented the utility of the collections by adding to the body of contextual data about them.

Academic research played an important role in the evolution of the modern nonprofit museum in the early twentieth century. Scholars and connoisseurs formed the basis of a class of professional museum workers. Curators, preparators, and conservators adopted codes of ethics and standards of practice that were instrumental in the development of museums and archives as educational institutions. However, until the 1950s, the primary audiences of both types of institutions were on-site visitors with specific, and often specialist, research needs rather than the casually curious.

During the past twenty years, a combination of changing professional attitudes, the interests of public and private funders, and the growing availability and reliability of reproduction technologies and electronic communication have resulted in a re-evaluation of museum and archival collections. The new target audience is the general education market and the new means of providing information to the target audience is electronic, most often via the Internet. The World Wide Web allows easy access to good quality image representations as well as to text-based contextual information about them. The public's expectation is that a broad range of information needs can and will be accomplished accurately via electronic surrogates without physical exposure to the primary sources from any place at any time. The worth of institutional assets is no longer gauged by looking at the collections inventory appraisal. It is now redefined as the combination of the physical materials in the collections, the

surrogates that satisfy a growing demand for visual information about them, and the text-based information that establishes their context and serves as the key to locating them.

Securing collections assets against misuse, theft, or damage is an ongoing concern of museums and archives. A variety of measures are implemented to safeguard collections. These include controlled access to storage and items on display, frequent inventories, environmental monitoring, administration of rights and releases, and strict procedures regarding use by staff members and others. Posting extra guards does not help to secure electronic information. And, unlike the Impressionist painting that is kept under surveillance or the Stradivarius violin that is rarely, if ever, removed from the display case, digital assets can be adversely affected by the very measures that are intended to ensure their integrity and authenticity. Security measures typically used in museums and archives to protect these assets are applied randomly at best and unintentionally at worst. Responsible stewardship of digital image assets calls for a more formal and thorough risk management assessment of potential threats and for the creation of an informed and thoughtful security plan for their management and protection.

Risk management is the sum of all activities directed toward acceptably accommodating the possibility of failure in a program. Risk management is based on assessment; every risk management assessment includes a number of tasks: (1) identification of concerns, (2) identification of risks, (3) evaluation of the risks as to likelihood and consequences, (4) assessment of options for accommodating the risks, (5) prioritization of risk management efforts, and (6) development of risk management plans (http://www.airtime.co.uk/users/wysywig/risk_1.htm). This article examines existing practices within museums and archives and provides suggestions on the creation of such plans as they apply specifically to the stewardship of digital assets.

DEFINING CONCERNS AND IDENTIFYING RISKS

Responsible individuals become concerned when a valuable possession is placed in jeopardy. The value of collections-related digital assets to museums and archives has been established. What are legitimate concerns regarding objects of value? Would these concerns be applicable to digital assets? It is possible to identify two obvious concerns. The first is fear that the asset itself will somehow lose value. The second is that the steward (in this case the institution and its professional staff) will somehow lose the asset or control over the asset.

How is value embodied in digital information and what would constitute a loss of value? The charter of museums and archives includes a mandate to preserve the information embodied in their collections. It seems reasonable to propose that the value of digital surrogates for

collections items lies in the relative ability of the surrogates to convey as much original information content as possible. The integrity of the digital image is judged as the degree to which it accurately represents its subject. If the information content of the surrogate is compromised, the surrogate is devalued.

There is a case to be made for the creation of very high quality, very high-resolution digital surrogates. These files are used as archival versions of image information, but reality intervenes when their content is put to practical use. High quality, high-resolution files are very large and therefore costly to store and transmit. The generally accepted rule is that the needs of different uses and users are best met by digital content presented in a variety of formats or resolutions, tailored to the situation. Accurate representation is in the eye of the beholder; the resolution and file size limitations dictated by intended Web use are not the same as those demanded by activities such as conservation assessment (Frey, 1997) (<http://lcweb2.loc.gov/ammem/formats.html>). As a result, every variant form of a digital file can and should be evaluated for integrity based on the use to which it is put.

Control over the asset is somewhat easier to describe and evaluate. The most obvious manifestation of control of image surrogates is the ownership of copyright and the ability to assign or to withhold assignment of use rights to others. There are other manifestations of control that are uniquely related to the museum or archive's responsibilities toward the public; these may in fact be more significant than copyright ownership. Nonprofit 501(c)3 charters and ethical responsibility dictate that it is not enough for institutions to own and care for objects. The legal definition of a museum includes the directive "to exhibit to the public on a regular basis" (Malaro, 1985). This has been interpreted for the last two decades as a mandate to educate by providing members of the public with meaningful and useful mediated access to collections. Control of the collections implies control of access to the collections in a proactive way. It is the job of museums to encourage and facilitate the use of collections and the information that they represent. Loss of control in this sense would mean an inability to effectively mediate the collections-related educational experience.

It is now possible to identify potential risks that are associated with each type of concern. Changes in the integrity of digital image information can be caused by direct modifications made to the image data. They may also be associated with modifications to contextual data that limit understanding and interpretation of the associated surrogate image. The destruction of the mechanisms that link contextual data to the appropriate digital image has the same effect as deleting contextual information. Loss of control over digital assets can be the result of failure to establish ownership and/or copyright. Even if copyright is established and

enforceable, failure to enforce rights has the same effect as having no rights at all. Failure to detect corruption of digital information means that invalid, partial, or inappropriate information will be spread under the guise of authentic reliable information. Each of these risks represents the possibility of an information systems failure.

PRIORITIZING RISKS—HOW SAFE IS STANDARD PRACTICE?

What are the chances that any of these risks will be realized? An examination of the typical ways in which digital image information and associated contextual data are created, managed, and made accessible sheds light on the probability of content degradation. Most institutions already employ both active and passive measures to prevent or minimize the impact of a reduction in reliable content in systems that depend on the use of digital image information. Do these efforts have any effect on the immediacy of each risk?

CREATING DIGITAL IMAGE FILES AND DERIVATIVES

A digital image cannot be a better representation than the best available from the conversion method used to create the image. A number of authors and research groups have conducted comparative studies of conversion techniques and produced recommendations for best-practice conversion methods, ranging from direct digital photography through microfilm and negative scanning to direct positive scanning and PhotoCD processing (<http://www.columbia.edu/acis/dl/imagespec.html>) (Kenney, 1997; Conway, 1996; Reilly, 1995). Similarly, the digitized image cannot be better than the source document or object without some sort of data modification. It is not necessary to belabor the importance of informed decision making in the process of creating archival image files from which derivative files may be drawn. Frey (1997) suggests that four targets be used for objectively evaluating the results of digitization: tone reproduction, color reproduction, detail and edge reproduction, and noise. Satisfactory performance in output tests of all four targets will guarantee that, at least at the archival level, an acceptably accurate digital representation of image information has been created.

The integrity of digital image information is inherent in the structure of the image file. Only bit-mapped images (those created from aggregations of discrete bits or units of data) are considered in this discussion; vector image data are created and used in museum and archival environments much less frequently than in academic libraries and special collections. The parameters that are chosen to define file structure determine the limitations of the file as an image surrogate. These parameters include dynamic range, resolution, and compression (Besser et al., 1995).

Dynamic range defines the ability of the file structure to convey tonal information about each pixel captured. Every digital image is composed

of a fixed number of pixels—tiny discrete blocks of tone. Bi-tonal images can only convey information in black and white. A bit (the basic building block of digital information) can only convey two possible values; therefore, bi-tonal information is conveyed using one bit per pixel. This type of information encoding produces the smallest possible files, but the resulting image cannot represent any range of shades between black and white. It is recommended for uses that involve modern printed works and line drawings or graphics and is frequently employed when the desired use is a printed reproduction of such materials. Gray scale uses 8 bits to represent each pixel, providing the capability of representing up to 256 shades ranging from pure white through gray to pure black. This format is usually recommended for representing black and white photographs, half-tone illustrations, and other two-dimensional representations that convey shading or variation in ink density. Color is best represented using 24 bits per pixel, which provides about 16 million different colors but which results in much larger file sizes. Color conveys much more information than gray-scale or bi-tonal files and is required for images in which color must be maintained but is also recommended for use in digitizing images of older documents (<http://www.columbia.edu/acis/dl/imagespec.html>; <http://lcweb2.loc.gov/ammem/pictel/index.html>). While software, printing, and display hardware designs determine the nature of the end product, the dynamic range of the image file establishes the bases from which these devices perform in tests of tone and color reproduction. Recording image data in a file structure that uses 8-bit color, for example, will in most cases result in image information that offers only a general approximation of the tonality of the original and severely impact the utility of the image surrogate for many uses.

Many institutions have chosen to protect their digital image assets by providing general access to only low-resolution files. Resolution refers to the number of pixels that are used to describe a single image (the fixed number mentioned earlier); it is usually expressed in terms of horizontal and vertical dimensions. An image recorded at a resolution of 512 x 768, for instance, has 512 rows and 768 columns of pixels. Resolution affects the level of detail that can be depicted by the image file. If a lower resolution is specified, fewer pixels will be used to describe the image and therefore edges may be blurred, areas of the displayed or printed image may appear blocky, tonal transitions may seem more abrupt, and detail may be lost altogether. An illustration will assist in visualizing the loss of information that may result from the use of lower resolutions.

Information conveyed by Figure 1, an extremely detailed photograph, would undoubtedly be lost if its digital surrogates were created using lower resolutions. Edge blurring would prevent a researcher from studying wheels, spokes, and hubs, and clothing detail would become invisible. The



Figure 1. Ford Auto Dealers Auto Parade, 1912 (neg. 0.3299). From the collections of the Henry Ford Museum & Greenfield Village.

wide range of tonal contrast across very limited spaces would also be obscured, and the overall effect would be a smoothing of shadows and features.

Compression is a technique used to reduce the size of a digital file. This is accomplished in a number of ways, including mathematical transformations and reduction of precision by the elimination of “noninformational” data or noise in the data set (Brown & Shepherd, 1995). Reduction of precision is the most commonly applied compression technique; the effect that it has on the quality of the resulting digital image makes compressed files very attractive to institutions that are concerned with potential unauthorized use of images. Some museums believe that, as in the case of employing low resolution, reducing the quality of the digital image file makes it unattractive to would-be electronic privateers. Reducing the size of a file means that the quality of the resulting image is reduced and that the amount of time that it takes to transmit the data from the file over a communications link is decreased. Using compressed files for Internet or intranet transfers therefore is doubly attractive, but there is a potential loss of image integrity that may prove significant depending on the use of the files. The risks are more obvious when compression algorithms are examined in greater detail.

The class of compressed formats termed "lossless" is based on data transformation algorithms (for a discussion of wavelet and fractal compression, see Puglia, 1998). In these formulas, the original scanned pixel values are transformed into other values, most often using either run-length encoding (i.e., Sunraster, TARGA, and TIFF format types 2 and 32773), LZW encoding (i.e., GIF LZW and TIFF scheme 5), or discrete cosine transforms, also known as DCT (i.e., JPEG DCT and MPEG DCT). One-dimensional differencing, a method employed to produce JPEG predictive implementation, a true lossless JPEG format, is not discussed here. In run-length encoding, repetitive sets of identical data values in the original data are replaced by codes, each made up of a single data value and a length value. The collection of codes and their values must be stored in the file as the "codebook." The resulting reduction in file size depends on the number of repetitive data sets in the original. The use of the compressed data is contingent on the ability of the user software application to use the codebook to decode the format.

Lempel-Ziv and Welch developed an alternative method (LZW) of encoding data in 1985 that also uses pattern recognition but allows the decoder to build the codebook as it processes the data stream. The resulting file is smaller than those created with run-length encoding. In formats that use DCT to compress files, a mathematical operation is applied to blocks of original data. The transformed block is represented by fewer bits in the digital file than the original block. Run-length and LZW encoding result in no loss of data; DCT does in fact result in some data loss due to round-off errors, but the overall effect on the quality of the resulting image is inconsequential. DCT tends to yield higher compression ratios; nevertheless, average ratios of original data file size to compressed file size tend to fall in a range of 2:1 to 9:1 (Brown & Shepherd, 1995, p. 190). TIFF and lossless or near-lossless JPEG formats are extremely attractive for the purposes of maintaining data integrity, but their application does not result in files that are as small as those created employing other "lossy" techniques.

Reducing the precision of data means eliminating information in the original file that is not necessary for the purpose at hand. The electronic scanning of a photograph, for example, may produce sets of greyscale values that, while different, are so close to one another in tonality that the human eye may not be able to distinguish a difference. Recording data at this level of precision is probably not necessary for the creation of an acceptable digital surrogate. Achieving more aggressive compression ratios, in the area of 20:1 and higher, requires the establishment of less stringent definitions of noise and results in more notable erosion of information content. Commonly used implementations of JPEG (there are twenty-nine total) use reduction of precision combined with other data encoding techniques to achieve compression ratios up to 100:1. In these

In Figure 3, the pocking of the glazed finish creates a uniform stippled pattern across the surface of the jug. On the jug, there is incised ornamentation in the ship that emphasizes the stippled effect. Compression of original image data scanned from this photograph would result in a reduction of tonal contrast across the jug and subsequent loss of fine detail in the resulting image surrogate.

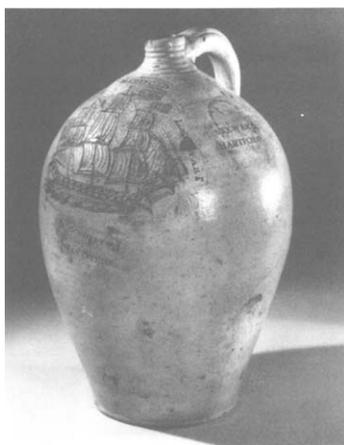


Figure 3. Salt-glazed stoneware jug, 1820–1830 (accession 57.67.8; neg. B.16963). From the collections of the Henry Ford Museum & Greenfield Village.

PURPOSEFUL MODIFICATION OF DIGITAL IMAGE INFORMATION

Museums and archives that were early adopters of digital image technology often discover that the electronic representations created in the first years of the digital revolution are less than satisfactory when compared to those produced with current technology. In the case of the Henry Ford Museum, an early version of the automated collections management system was designed to work in tandem with laser disk readers. Laser disk images were created from 90,000 photographs from video documentation. Now transferred to PhotoCD, even the good images (degraded by processing three steps removed from the original) are difficult to use without some manipulation. Color modification is the most obvious intervention that is applied to older digital image files. Commercial image manipulation software packages provide a variety of other techniques to modify file information. At times there are side effects caused by the image processing operations that are used to enhance a problematic image in the form of modifications of image information that does not directly relate to the condition being corrected. For example, noise suppression using a method called Gaussian smoothing often results in the blurring of edges on shapes within the image (Davies, 1997, p. 44).

Furthermore, image enhancement operations that result in changes of brightness or contrast, noise reduction, and the sharpening of edges may employ filtering and thresholding algorithms that cause edges to shift in position and image shapes to become distorted. Curves and circles are particularly susceptible to shift. In images that contain both straight and curved edges, shapes may appear to move in relationship to one another after noise suppression filtering is applied (p. 59).

Attempting to correct noise or modify contrast in an image based on data scanned from Figure 4, a photograph of the Ford Rouge Plant, could result in distortion of shapes and subtle changes in the perspective of the image. If this occurred, the digital representation would present a false picture of the location of the camera, the size of components, and their spatial relationship to each other.



Figure 4. Conveyor System and Power Plant Stacks, Ford Motor Company Rouge Plant, 1927. Photograph by Charles Sheeler (neg. B.189.6577). From the collections of the Henry Ford Museum & Greenfield Village.

Handwritten letters, old photos, and artifacts are not the only collections items documented digitally in museums and archives. Figure 5, an image of a Model A parts drawing, was printed from an image in a collection of large-format microfilms, the only existing copies of this and other drawings. The original drawings and production copies were destroyed; the microfilm is the only remaining resource for specifications used in the reproduction of authentic parts. Any edge shifts or spatial distortions caused by manipulations of the digitized versions of these drawings could lead to disastrous misinterpretation of the images by the parts manufacturers who use them.

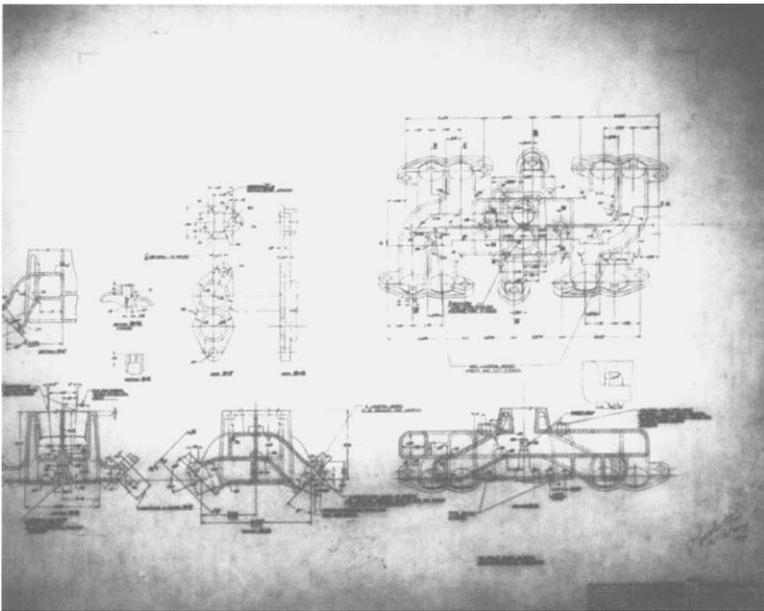


Figure 5. Model A Parts Drawing, 1944. From the collections of the Henry Ford Museum & Greenfield Village.

Most purposeful image data modifications cause the existing image characteristics to change or disappear altogether. A proactive security measure taken by some institutions to protect ownership is the addition of digitized credit line information that is either superimposed on or appended to the original image data. The thumbnail images displayed in Just In Time Images, the photo reproduction page on the Henry Ford Museum & Greenfield Village Web site, are altered in this fashion (<http://www.hfmgv.org/jit/still/index.htm>). The overlaid information obviously takes the place of the image data that formerly occupied that space. In the event that credit information is appended to an existing

image, the addition of pixels to the existing file results in the deletion of pixels from some other location in the image. The usual result is a cropped image. Electronic cropping may also be a purposeful action, prompted by display size limitations or aesthetic considerations. Regardless of reason, the elimination of digital information results in changed orientation of the image elements to the boundaries of the image.

The measuring rule normally included in documentation photos has been cropped out of this image of a single artifact (Figure 6). As a result, there is no referential information to provide a sense of the scale of the subject. Cropping can also remove spatial reference points that typically occur near the edges of images such as horizon lines.



Figure 6. Painted tin box, ca. 1825 (accession 29.737.2; neg. B.62832). From the collections of the Henry Ford Museum & Greenfield Village.

Failure to detect corruption of digital image information means that invalid, partial, or inappropriate information will be spread under the guise of authentic reliable information. It is important to re-emphasize that variation in data integrity among multiple surrogate versions of the same image is acceptable because of storage size, level of detail, and delivery speed requirements imposed by different uses.

THE ROLE OF METADATA IN MAINTAINING IMAGE INTEGRITY

It can be argued that any information lost as the conscious or unconscious result of the processes described earlier could be restored intellectually if the digital image information is associated in some way with contextual metadata. Museum collections management systems usually display thumbnail images on the same screen as catalog information for the artifacts that are documented by those images. The text descriptions of the artifacts, if sufficiently rich and detailed, assist in the interpretation of the image information and vice versa. Image file technical metadata assists

in the evaluation of the digitized image based on the nature and limitations of its method of creation. Yet the two sets of complimentary information, metadata and digitized image, are usually stored in separate files. The catalog data exist in a series of rows linked across the tables of a relational database by virtue of unique identification numbers. The image data are stored in discrete files, one for each digital representation. The names of the files often bear no resemblance to the name or the accession ID of the original object. The connection between the sets is a one-way street leading from the text data to the digital image file. It is also fragile; the loss or modification of data in a single image file name field prevents the user of the catalog from viewing the image as well as preventing the user of the image from viewing associated catalog entries. If the connection is somehow broken and if the sets represent hundreds of thousands of artifacts or manuscripts, it will be almost impossible to properly relink all of the records and files. Standards related to the composition of contextual metadata aside, there is a serious need to consider the adoption of image data file formats that in some way automatically incorporate metadata in their structure. There are numerous informative discussions on the topic of metadata in museum and archival applications available both in print and on the World Wide Web (<http://www.cimi.org>; <http://www.gi.getty.edu/index/warwick.html>; <http://www.acctbief.org/avenir/images.htm>).

OWNERSHIP, COPYRIGHT, AND CONTROL OF DIGITAL IMAGE ASSETS

In a world of increasingly complex legal issues, few pose more varied and vexing problems than those surrounding copyright and the ownership of images and image surrogates. Copyright laws were created to protect the rights of individuals to own the expressions of their ideas (Malaro, 1985, p. 113). Copyright is actually a suite of rights that may be conveyed, transferred, or retained, singly or in sets. Copyrights include (a) the right to reproduce the work, (b) the right to produce derivative works from the original, (c) the right to distribute copies for sale, (d) the right of performance, and (e) the right to display the work. Before 1978 in the United States, copyright existed only if the artist distributed the work with the copyright symbol; failure to do so was deemed a waiver of copyright. Copyrights to works acquired by a museum were assumed to transfer to the museum unless specific statements were made to the contrary. After the revision that took effect in January 1978, copyright was considered implicit in the act of creation and could only be waived by a statement to that effect. Museums can no longer assume that rights transfer automatically.

Until recently, expression of ideas implied the act of creating something with physical presence: a book, a painting, or a better mousetrap. Rights of authorship could not be enforced without recourse to referencing

something tangible or a tangible copy of a work. Digital representation is not easily categorized as having physical presence; there is no question that original work is involved, but marking or branding or seizing control of the "thing" that is created either as an original work or copy is conceptually difficult. As John Barlow (1996) describes the situation, under original copyright law "the bottle was protected but not the wine. Now the bottles are vanishing" (p. 11). Digital assets are the wine without the bottle. Controlling the use of digital image information representing items to which the museum clearly has copyright is difficult due to the accuracy with which duplicates can be made and the speed with which they can be disseminated (Bearman & Trant, 1997). Ambiguity regarding rights to photographic and digital reproductions of works in the public domain further complicates the process of enforcement and control (Akiyama, 1997). These reproduction rights, historically defended by museums and used to generate licensing income, have been threatened by a recent court decision that has implications for the control and use of digital reproductions. In this case, the Bridgeman Art Library, a British company that licenses transparencies of public domain art works that are owned by museums and collectors, brought suit against Corel Corporation, makers of a CD-ROM product containing digital reproductions of well-known paintings including 120 from the Bridgeman portfolio. Corel neither licensed nor requested permission from Bridgeman to use the works over which Bridgeman claims to have sole control. Bridgeman maintained that Corel had violated their copyrights; Corel countered by claiming that the museums and collectors could not assign to Bridgeman the rights that pertain to works in the public domain. The court ruled in favor of Corel, finding that substantially exact photographic reproductions of two-dimensional works of art are not copyrightable because they do not involve original work ("Copyright Case," 1999). The implications of this decision are serious. If it is upheld, museums will neither be able to exercise control over the use of image reproductions of public domain items in their collections nor to charge copyright fees for the use of such images, no matter the format.

Assuming that an institution's rights to the digital representation of an image are established, copyright enforcement can be accomplished in two ways. The institution can control access to the digital file permitting use only by those who are appropriately authorized. On the other hand, the museum or archive can provide unlimited access to digital visual resources that are marked with an indication of proper ownership. Suspect reproductions can then be examined and, if the mark is detected, the institution can proceed with steps to enforce their rights.

Most collections management software packages provide users with one or more security schemes. These are implemented by logon id and selectively allow each user to perform pre-defined sets of operations on

specific fields or files. Application system security can be an effective way to prevent the modification of text-based contextual information. Digital envelopes can also be used to protect text files that contain metadata relating to digital image files. A digital envelope uses encryption to permit access to file content on a selective basis. The text data are encrypted using a key and then the key itself is encrypted using another key. The user must decode the key data before it can be applied to the content in a second decoding step; double keys protect the content from both casual theft and from most true data pirates.

Digital image information is stored in discrete files separate from a text-finding aid or catalog information. The image files are accessible to proprietary system users and to everyone else with image server access as well. The most recent implementations of Microsoft Windows provide image display capabilities as default readers that respond to Open commands; the user can invoke them by selecting the image file name from any storage device. These files can be modified using any commercial image processing software. While it is possible to store image file information in a digital envelope (Acken, 1998), this technique requires that all potential legitimate users be identified and equipped with appropriate keys. It is often undesirable (as in the case of images used on a Web site) to prevent casual viewers from seeing an image. In this case, marking the image files and monitoring their use is an effective way to protect content and enforce copyrights if necessary.

Digital watermarking is the process of inserting marks or labels into digital content in such a way that they are unobtrusive yet inseparable from the source data (Yeung, 1998, p. 32). This article has already referenced the use of visible credit lines superimposed over the source image. This technique visibly alters the content of the surrogate image, displacing potentially meaningful data. Most digital watermarks are transparent. There is no degradation of visible content caused by the watermark, but the watermark is detectable using special software processes. A good analogy can be drawn from photocopying. At the Henry Ford Museum Research Center, copies of photographs or documents from the collections are made on a photocopying machine using paper that is pre-stamped with a rights and use warning in red ink. The red message displaces meaningful data from the source document. If watermarked bond paper without the stamp was used for the copies, no meaningful source data would be displaced but the watermark could still be viewed under certain circumstances, as on a light table.

The analogy breaks down when a reproduction of the photocopy on bond paper is made. The watermark will not appear as part of the photocopied information, although the overall quality of the content will degrade as copies are made from other copies. In digital watermarking, the file contents can be duplicated an infinite number of times with no

degradation of quality, and theoretically the watermark will appear the same in every copy.

Watermarking technology is opportunistic, relying on the fact that in any digitized image file (including compressed files) there are some bits that carry less significant information than other bits. In invisible watermarking, modification of the data in these bits causes minimal visible change in the image when displayed or printed (Memon & Wong, 1998). The modifications are data substitutions that collectively make up the watermark. The degree to which the discernible image content is affected depends on the nature of the image (if it contains large areas of solid intense color for example) and the nature of the watermarking algorithm (Wayner, 1997). Visible watermarks affect the image, usually by adding a transparent logo or visual message to the displayed data. In both types of watermarking, modified data are located in different places or "holes" in the image file and can be extracted and assembled to convey meaningful information.

There are ways to remove watermarks but benchmark standards for robustly resistant watermarks are being developed (Mintzer et al., 1998). Robust watermarks are those which can be recovered in spite of intentional or unintentional modification of the image file. They must be able to survive a variety of processes including filtering, cropping, scaling, and compression. This type of watermark is useful for establishing ownership of an image or for detecting unauthorized copies. There is another form, fragile watermarking, that relies on the ability of the mark to break easily if the image is altered. Fragile watermarks are designed as tools for identifying compromised data; they can even shed light on the nature of the alteration. If the ruling in the *Bridgeman vs. Corel* case is not struck down, fragile watermarking may be the only way to ensure that uncontrolled use of digital files does not result in a proliferation of inaccurate and unauthentic images on the World Wide Web.

DEVELOPING A DIGITAL IMAGE RISK MANAGEMENT PLAN

Risk management plans should be developed based on the unique nature of each institution's digital image holdings and the audiences that access them. It is useful to recall that the purpose of risk assessment is to develop acceptable accommodations of failure. Perfection is neither obtainable nor necessary. Few, if any, institutions have the resources to create, store, and use images that are perfect electronic replicas of the originals by current standards. It is enough to be aware of the compromises that are made and of the impact that they may have now and in the future. It is also important to be informed and open to change.

There are a number of development efforts underway that could have a major impact on the manner in which museums and archives use and distribute digital image information. One example of an exciting emerging

technology is the FlashPix image file format, developed by a consortium of high-tech companies including Eastman Kodak, Hewlett-Packard, Live Picture, Inc., and Microsoft (Donovan, 1998). FlashPix addresses a number of problems. It allows the storage of original digital input plus a number of lower resolution copies in the same file. Each resolution is broken into smaller segments called tiles that can be read individually or in groups. FlashPix also allows text metadata to be stored in the same file as the image data, solving the problem of developing standardized headers or maintaining links between image files and contextual data stored elsewhere. Although not currently supported in browser software, this format could greatly simplify the digital image risk management process.

Digital watermarking technologies are also changing rapidly. Commercial applications are concentrating on rights enforcement and signature authentication applications, but there is a growing interest in using the "holes" in digital image files for the storage of metadata. One author suggests that embedded hyperlinks could direct viewers to related Web sites and that embedded indexing data could be used to pre-select images for viewing (Acken, 1998 p. 77).

There is an obvious need for ongoing research, evaluation, and planning if museums and archives are committed to protecting their digital image assets. A number of potential threats to the integrity of digital image information have been identified here. Changes in the integrity of digital image information can be caused by the manner in which the source data are acquired and recorded and by modifications made to the image data file. Alterations made to contextual data can limit valid interpretation of the associated surrogate image. The destruction of the mechanisms that link contextual data to the appropriate digital image has the same effect as deleting contextual information. Loss of control over digital assets can be the result of failure or inability to establish and publicize copyright. Even if copyright is established and enforceable, failure to enforce rights has the same effect as having no rights at all. Finally, failure to detect corruption of digital information means that invalid, partial, or inappropriate information will be spread under the guise of authentic reliable information. Some institutions are already proactively applying security measures to digital image collections. As noted here, security measures can have a negative impact on the integrity of the files that they are designed to protect. Systematic consideration of risk factors can inform the creation of procedures and application of security that works to guarantee the reliability and accuracy of digital image assets.

REFERENCES

- Acken, J. M. (1998). How watermarking adds value to digital content. *Communications of the ACM*, 41(7), 74-77.
- Akiyama, K. A. (1997). Rights and responsibilities in the digital age. *Visual Resources*, 12(3-4), 261-268.

- Barlow, J. P. (1996). Selling wine without bottles: The economy of mind on the global net. In P. Ludlow (Ed.), *High noon on the electronic frontier* (pp.1-8). Cambridge, MA: MIT Press.
- Bearman, D., & Trant, J. (1997). Museums and intellectual property: Rethinking rights management for a digital world. *Visual Resources*, 12(3-4), 269-280.
- Besser, H., & Trant, J. (1995). *Introduction to imaging: Issues in constructing an image database*. Santa Monica, CA: Getty Art History Information Program.
- Bridgeman copyright case update. (1999). *Aviso* (April), 1.
- Brown, C. W., & Shepherd, B. J. (1995). *Graphics file formats: Reference and guide*. Greenwich, CT: Manning Publications.
- Conway, P. (1996). *Conversion of microfilm to digital imagery: A demonstration project: Performance report on the production conversion phase of Project Open Book*. New Haven, CT: Yale University Library.
- Copyright case challenges long-held museum assumption. (1999). *Aviso* (February), 1.
- Davies, E. R. (1997). *Machine vision: Theory, algorithms, practicalities*. San Diego, CA: Academic Press.
- Donovan, K. (1998). The promise of the FlashPix image file format. *RLG Diginews*, 2(2). Retrieved October 1, 1999 from the World Wide Web: <http://www.rlg.org/preserv/diginews/diginews22.html#FlashPix>.
- Frey, F. (1997). Digital imaging for photographic collections: Foundations for technical standards. *RLG Diginews*, 1(3). Retrieved October 1, 1999 from the World Wide Web: <http://www.rlg.org/preserv/diginews/diginews3.html#com>.
- Kenney, A. R. (1997). The Cornell Digital to Microfilm Conversion Project: Final report to NEH. *RLG Diginews*, 1(2). Retrieved October 1, 1999 from the World Wide Web: <http://www.rlg.org/preserv/diginews/diginews2.html#com>.
- Malaro, M. C. (1985). *A legal primer on managing museum collections*. Washington, DC: Smithsonian Institution Press.
- Memon, N., & Wong, P. W. (1998). Protecting digital media content. *Communications of the ACM*, 41(7), 35-43.
- Mintzer, F.; Braudaway, G. W.; & Bell, A. E. (1998). Opportunities for watermarking standards. *Communications of the ACM*, 41(7), 56-64.
- Puglia, S. (1998). Fractal and wavelet compression. *RLG Diginews*, 2(3). Retrieved October 1, 1999 from the World Wide Web: <http://www.rlg.org/preserv/diginews/diginews23.html#technical2>.
- Reilly, J. M. (1995). Technical choices in digital imaging: The Technical Images Test Project in review. In P. McClung (Ed.), *RLG Digital Image Access Project* (Proceedings from an RLG symposium held March 31 and April 1, 1995, Palo Alto, CA) (pp. 85-93). Mountain View, CA: Research Libraries Group.
- Wayner, P. (1997). *Digital copyright protection*. Boston, MA: AP Professional.
- Yeung, M. M. (1998). Digital watermarking. *Communications of the ACM*, 41(7), 30-33.

ADDITIONAL REFERENCES

- CNRI Registry. (1998). *Handle systems overview*. Retrieved October 1, 1999 from the World Wide Web: <http://www.handle.net/overviews/hs-version4.html>.
- Consortium for the Computer Interchange of Museum Information. (1999). *Home page*. Retrieved October 1, 1999 from the World Wide Web: <http://www.cimi.org>.
- Getty Information Institute. (1997). *Metadata standards*. Retrieved October 12, 1999 from the World Wide Web: <http://www.getty.edu/gri/standard>.
- Image Quality Working Group of ArchivesCom. (1997). *Technical recommendations for digital imaging projects*. Retrieved October 1, 1999 from the World Wide Web: <http://www.columbia.edu/acis/dl/imagespec.html>.
- Library of Congress Preservation Office. (1998). *Manuscript digitization demonstration project final report*. Retrieved October 12, 1999 from the World Wide Web: <http://memory.loc.gov/ammem/pictel/index.html>.
- National Digital Library Program, Library of Congress. (1998). *Digital formats for content reproductions*. Washington, DC: C. Fleischhauer. Retrieved October 12, 1999 from the World Wide Web: <http://memory.loc.gov/ammem/formats.html>.

- Sandore, B. (1997). Images and their descriptive metadata. In *Proceedings of the Conference on the Future of Communication Formats* (Held in Ottawa, Canada, October 5-10, 1996, sponsored by the Banque Internationale des Etats Francophones and the National Library of Canada) (pp. 121-133). Ottawa, Ontario, Canada: Banque Internationale des Etats Francophones.
- Simmons, C. (1998). *Risk management*. Retrieved October 1, 1999 from the World Wide Web: http://www.airtime.co.uk/users/wysywig/risk_1.htm.