Wisconsin Heritage Online Preservation Recommendations

Prepared by the WHO Preservation Working Group, June 2006
Revised and updated August 2009

The purpose of this document is to help Wisconsin Heritage Online Content Providers understand the technical requirements of a digital preservation program so that they can make informed decisions and understand the implications of their choices for the future. The document includes a "pyramid" of possible preservation activities stratified by level of complexity and cost. All Content Providers are encouraged to develop a program utilizing those levels of the pyramid they are most comfortable with. Additional information is also provided to better understand current thinking and activities in the preservation of digital research assets.

Members of the 2006 WHO Preservation Working Group
Joshua Ranger (chair) – UW Oshkosh
Marlys Brunsting – UW Green Bay
Steven Dast – UW Madison
Anita Doering – La Crosse Public Library
David Null – UW Madison
Ed Van Gemert – UW Madison
David Weinburg-Kinsey – Cardinal Stritch University

Introduction

Converting materials into a digital format is not a preservation activity. However, that is not to say that there may not be long-term preservation benefits to digitizing material. For example, creating copies from the digital image rather than the original material results in decreased handling and may lengthen the life of the original material. The Society of American Archivists suggests scanning original materials once and into a non-proprietary file format so that damage from scanning is kept to a minimum and yet the image can be used many times and eventually migrated into new formats as necessary.[1]

Throughout this document, we are talking about digital preservation rather than the preservation of the original material. Many issues arise when discussing digital preservation, such as the longevity of the media on which the information is stored and the ongoing obsolescence of computer hardware and software to read the information. Digital files must be maintained, backed up, refreshed, and migrated on a regular, ongoing basis to remain accessible with current hardware and software. For these reasons, the act of digitization is still not considered a preservation standard, although standards do exist for digitization practices.[2]

It is the responsibility of the creators and keepers of Wisconsin's digital assets to take steps to ensure that their digital files will remain accessible over time. This requires an organizational commitment to the ongoing maintenance of the digital collection and any resulting costs, including maintaining server security and ensuring that the restoration of applications and data from backups is always possible.

---

[1] For a good example of "best practices," see "BCR's CDP Digital Imaging Best Practices, Version 2.0," June 2008: http://www.bcr.org/cdp/best/digital-imaging-bp.pdf

[2] Sammie L. Morris, "Preservation Considerations for Digitization of Archival Materials," *Archival Outlook*, May/June 2005, p. 9; 26.

Section I: Preservation Activity Options (Preservation Pyramid)

"Digitization must not be viewed as a substitute for other preservation activities. For those responsible for the longevity of collection materials in an age of technological flux and uncertainty…valuable materials, such as rare and unique archival documents, photographs, and artifacts, must be retained in original form regardless of the existence of digital or microfilm surrogates…the permanent accessibility of the electronic information is of utmost concern"[3]

Wisconsin Heritage Online is a cooperative venture among diverse cultural heritage institutions. The wide-ranging needs and institutional realities of these diverse organizations require a varied set of options in the preservation of digital assets. In addition, different projects from a single institution may have different needs, and an institution may choose to follow a different course for one project over another. Some questions to ask when developing a preservation plan include: How long do we intend to make this project available electronically? Do we wish to avoid rescanning this material in the future? What are the possible uses and users of this material that we may not have considered?[4]

Much more so than in the realm of stable media, such as paper and microfilm, the preservation of digital resources is dependent on regular maintenance of the media and files in which the resources are housed. As such, the question of preservation is fundamentally one of convincing our future selves and successors to engage in the activities necessary for the survival of the digital assets we create. One might reasonably suppose that such preservation will occur as long as the value of the digital assets exceeds the cost required to maintain them.

Our preservation recommendations, therefore, are based on maximizing the value of the digital resources created while minimizing the probable future costs for maintenance. Of course, this must also be balanced against the present costs of creating, processing and storing the digital material. As a result, we propose a multi-tiered set of recommendations for each work stage to allow individual participating institutions to evaluate and select a digitization strategy which best balances their own available resources and capabilities with their assessment of future needs and value.

Preservation Recommendations

Note: Preservation levels in one area are generally not related to those in any other area, so institutions are free to mix and match options as desired. Nevertheless, Level 0 in each area is considered to be below preservation quality, negating any preservation actions taken in other areas. *Therefore, an institution intending to*

---

[3] Morris.

[4] There are many books available that address such questions and most digital conferences and workshops cover these issues. For example, Arizona State Library, Archives & Records, *Digital Projects Guidelines* (Phoenix, AZ), 2003.

create long-term digital resources will need to follow at least the level 1 recommendation in each area.

Scanning

Decisions made at the point of scanning primarily represent a trade-off between the file size of the resulting image and the amount of information it contains. At present, storage costs are a genuine concern and preclude a strategy of simply scanning everything at the highest resolution available. Current trends suggest that in the future storage will be plentiful and inexpensive. The critical issue, then, is whether the digital images created today have sufficient information to continue to be valuable into the future.

Level 0 – Scan materials at a resolution that equals or exceeds (by 1.5 - 2x) the largest size that will be used for online display. Choose a bit-depth and color mode appropriate to the material (1-bit bitonal, 8-bit grayscale, or 24-bit color).

Present ramifications: Storage requirements for material are minimized; fine print, particularly on maps, may be too small to read even on master image.

Future ramifications: Resolution of images will be insufficient for other purposes (such as printed reproductions or zooming to examine detail) or advancing technology (high resolution displays) and will have to be rescanned.

Level 1 – For photographic prints, scan at 300 dpi if the longest edge of the item is 7 inches (18 cm) or longer. Scan at 600 dpi if the longest edge of the item is shorter than 7 inches (18 cm). Use 8-bit grayscale (= 256 grays) or 24-bit color (= millions of colors). Scaling, if available, should be set to 100%

For photographic film (slides, negatives), scan 35 mm film at 2000 dpi. For other sizes of film, set resolution to produce a 3000 pixel image along the longest dimension. (e.g. for a 4″ x 5″ negative, scan at 600 dpi: 600 dpi x 5″ = 3000 pixels) Other settings should be as for photographic prints

For printed text without illustrations, or simple line art only, scan at 600 dpi. Use 1-bit (= bitonal). Take care to adjust threshold (brightness) so text is clear with well-formed letters (minimizing break-up or fill-in) without introducing extraneous noise. If text is otherwise clean, darker, bolder text is preferable to lighter text.

For text with color, grayscale, or fine line illustrations; hand-written or poorly printed text; and text on stained or discolored paper, scan at 300 dpi. Use 8-bit grayscale (= 256 grays) or 24-bit color (= millions of colors).

Present ramifications: Storage requirements may be significant, particularly for large numbers of scans (estimate 2-10 MB per grayscale image, 6-36 MB per color image).

Future ramifications: UWDCC (University of Wisconsin Digital Collections Center) guidelines are based on broadly accepted library and archive recommendations which are themselves based on a theory of full information capture--that is, the specified resolution is matched to the level of detail found in the original object in order that the digital version will serve to convey the same information that the original was intended to convey. This theory is most applicable to the realm of textual material, specifically material in roman characters, but is reasonably sound for graphical material as well. Since the full information of the original is captured, there should be no question of future obsolescence; the guidelines also correspond well with specifications for print reproductions. Researchers interested in artifactual details may be unsatisfied by these images, but they are likely to be very much in the minority and may want to examine the originals in any event.

Level 2 - As level 1, but include a grayscale or color target in the image for all grayscale or color scans.

Present ramifications: Storage requirements increase slightly over Level 1 as the scan area must be increased to capture the targets.

Future ramifications: May serve to increase visual fidelity and/or user confidence in the image. Original characteristics more easily reconstructed if the appearance has been altered in the scanning or image enhancement process. Only beneficial if the targets are retained in the archival copy or are used by a trained individual in the process of image enhancement (see below).

Level 3 - As level 2, but scan textual material as grayscale (do not use bitonal scanning) and scan all material at a minimum of 400 dpi. Scan material from the earliest generation possible (e.g. photographic negatives instead of prints, original text pages instead of photocopies).

Present ramifications: Storage requirements increase for all images, quite significantly for material that would have been scanned bitonally.

Future ramifications: Greater fidelity and detail increase the likelihood that the images will suffice for those users interested in artifactual features.

Level 4 - As Level 3, but scan all photographic material as color, use 16-bit grayscale and 48-bit color (instead of 8-bit and 24-bit). If possible, use a fully color calibrated workstation for scanning.

Present ramifications: Storage requirements double for all material; increase six times for black and white photographs.

Future ramifications: Very subtle gradations of shade and color that would otherwise be discarded are retained. This may aid in computer-assisted analysis of material such as bringing out hidden detail. Only beneficial if the original scan is retained as an archival copy (Image Enhancement Level 3 or 4)

Image Enhancement

Decisions made during the process of image enhancement represent a trade-off between fidelity to the original item and quality of presentation. With very few exceptions, any image enhancement operation introduces some mathematical error or loss of information into the data that describes the images. Generally, this loss is visually insignificant, but it could potentially interfere with future use of the images. Also, depending on the processes used, it may be impossible to reconstruct the appearance of the original item from the processed image, compromising its value as a surrogate object.

On the other hand, avoiding image enhancement entirely is likely to lead to an objectionable, “sloppy” appearance in the online environment. Skewed images are particularly noticeable in the context of perfectly rectangular computer screens and application windows. Dark, muddy, or washed-out photographs will not appeal to users and may be difficult to pull information from, however true to the original’s appearance they might be.

Level 0 - Use any image enhancement desired.

Present ramifications: Presentation image can be made to look as pleasing as desired (though this is largely up to the judgment of the operator). However, image enhancement costs may be significant without clear boundaries as to what processes should be used.

Future ramifications: Loss of fidelity to the original may be significant. It will likely be difficult to reconstruct with accuracy certain aspects of the original item.

Level 1 - Restrict image enhancement to cropping, rotation (de-skewing), levels or curves, and basic color correction.

Present ramifications: The most obvious visual distractions can be corrected so that the online appearance is pleasing.

Future ramifications: Information loss from enhancement may still be a concern; an exact reconstruction of the original scan is not possible.

Level 2 - Restrict image enhancement to cropping (without rotation), rotation in 90 degree increments.

Present ramifications: Sloppy appearance for at least some images. But, costs for image enhancement are minimized

Future ramifications: No mathematical error is introduced by these operations, so the original scan is, in essence, retained. Any information cropped off is, of course, lost, so care should be taken with cropping.

Level 3 - Use the unenhanced original scan as your archival copy. Use any image enhancement desired on derivative master.

Present ramifications: Presentation image can be made to look as pleasing as desired. Uncropped image requires somewhat more storage space than if cropped. Image enhancement costs may be significant.

Future ramifications: Original scan is retained without information loss from image manipulation or over-enthusiastic cropping. However, the results of the image enhancement process are not preserved so may need to be reconstructed in the future if the distribution images are somehow lost.

Level 4 - Save both a copy of the unenhanced original scan and the enhanced derivative master as your archival copies. Restrict enhancement operations to those listed under Level 1 only.

Present ramifications: The most obvious visual distractions can be corrected. Storage requirements are doubled. Some form of version control, i.e. through file naming conventions, is needed.

Future ramifications: Availability of both files maximizes flexibility of future use and re-use of images. Future maintenance may be complicated by the existence of multiple versions.

File Format

The availability and usability of a particular file format is critically dependent on software that can encode and decode files in that particular format. One of the greatest dangers to the preservation of digital assets is format obsolescence--that is, the lack of software that can be used to decode the files in which the assets are stored. More basic formats are easier to write software for and more likely to broadly supported, but may lack desirable features of more advanced formats. In our case the most likely desirable feature is that of storage efficiency, i.e. data compression. Therefore, the choice of file format (and compression) represents a balance between present storage costs and future maintenance costs.

Level 0 - Save all preservation images in a non-proprietary standards-based format (e.g. TIFF, JPEG, PNG, JPEG2000); minimize use of lossy compression (e.g. set JPEG quality to high or maximum)

Present ramifications: Possibility of some storage savings from compression

Future ramifications: Standards-based formats maximize the likelihood that future users will have software (or be able to construct software) to decode and view the images. Using multiple formats, however, increases the level of complexity of maintaining the images, and increases the number of occurrences of format-obsolescence that will have to be handled. Furthermore, lossy compression, like image enhancement, results in the loss of (in theory, visually insignificant) information that cannot be reconstructed, especially when the loss is reiterated over several cycles of opening and resaving.

Level 1 - Save all preservation images in TIFF format using no or lossless compression.

Present ramifications: Storage savings reduced but still possible

Future ramifications: Standardizing on a single, currently ubiquitous standard simplifies future maintenance. However, broad nature of the standard means that compression schemes that are currently allowed in the standard and supported by some software may not be universally supported and may be dropped in the future. Use of lossless compression ensures that the preservation image can be reconstructed exactly without loss of information.

Level 2 - As Level 1, but avoid compression for color and grayscale images; bitonal images may use CCITT-Group IV compression.

Present ramifications: Storage savings further reduced, although savings from bitonal compression are still significant.

Future ramifications: Compatibility and obsolescence problems with varying compression options are largely avoided, simplifying future maintenance. Current broad use of Group IV bitonal compression scheme provides an ample cushion of safety, but adds an additional area of complexity and potential obsolescence.

Level 3 - As Level 2, but avoid compression even for bitonal images.

Present ramifications: No storage savings

Future ramifications: Simplest possible scenario for future maintenance

Physical Format

Information is only as stable as the media on which it is stored. Digital information is particularly fragile for at least two reasons. First, it is much less robust than information stored on analog data when it comes to handling a partial loss of information--a digital file is likely to lose its ability to be opened at all or garbled when enough errors are encountered. Second, digital information is generally stored in a very physically compact way so that minor physical damage or degradation can result in the loss of a large amount of data, while a serious physical disaster can cause the loss of a tremendous amount of data.

Strategies for defending against problems with physical media primarily involve a combination of redundancy and monitoring to detect degradation before it becomes severe. The choice of physical format also affects the ease and speed with which the files can be accessed when needed, and the ability to add storage capacity as needed.

Level 0 - Save preservation images onto a local hard drive or LAN and pray.

Present ramifications: Easiest, least expensive option (since you had to buy a computer to scan anyway). Instant access to files, but additional storage capacity may be difficult to add.

Future ramifications: At some point, your hard drive will crash and the files will be permanently lost.

Level 1 - Burn preservation images to duplicate (ISO 9660) CDs with post-burn verification, i.e., burn preservation copies to CDs using software such as Roxio.

Present ramifications: Hardware and media are relatively inexpensive. Operation can be performed on a standard desktop computer. Physical storage must be identified for the CDs. Files can be accessed fairly readily, depending on where the CDs are stored.

Future ramifications: CDs are generally stable, though not universally so, and are vulnerable to damage from improper handling and storage. Because CDs are stored on the shelf, outside of a computer network, monitoring for problems is slow and inefficient. Redundancy helps to protect against some problems, but problems could develop on both copies, particularly if they are not closely monitored. CD format is well established and unlikely to become obsolete suddenly. A long refreshment and migration cycle is possible (as much as every 10-20 years), but increases the chances of data loss.

Level 2 - As Level 1, but also implement a data-integrity checking program for the CD (based on, for example, checksum files included on the CD).

Present ramifications: As Level 1, but with increased expense and labor for setting up the data checking system and additional processing when burning CDs

Future ramifications: Checking CDs for problems can be automated to some extent, but still requires physical handing of the media and an active effort to initiate the checks.

Level 3 - Back up preservation images to duplicate (or triplicate) magnetic tapes, with a provision for archival storage in multiple locations.

Present ramifications: Hardware and media can be expensive (but reasonable given their capacity), and may require more advanced technical support to install and operate. Physical storage is greatly reduced compared with CDs, but remote storage must be arranged. Access to files may be slow (hours or days to retrieve a file) depending on where media and hardware are housed and operated.

Future ramifications: Tape media may be less stable in the long-term than CDs, but is probably less vulnerable to short-term damage. Tape generally includes robust error correction mechanisms to help minimize the effect of minor degradation. Remote storage defends against disasters in any one location. However, tape formats are continuing to evolve, so obsolescence is a significant concern. Refreshment and migration needs to be undertaken regularly (perhaps every 3-5 years).

Level 4 - As level 3, but also store preservation images on a RAID device.

Present ramifications: Hardware is significantly expensive and requires dedicated technical support. Enables instant access to files.

Future ramifications: Live access to files enables possibility of automatic integrity checking and simplifies process for mass migration of files. Media is built in to hardware, minimizing risk of handling damage. Redundancy of information across disks enables automatic recovery from individual disk failures; tapes enable recovery from more extensive hardware failures. Tape refreshment and migration is still required, but availability of files on RAID simplifies process.

Metadata

The metadata should be in a non-proprietary, standards based format. Generally, this means exporting the data into some ASCII-based file. However, the real value in metadata is its organization and interpretation, so the metadata needs to be based on a standard scheme, like Dublin Core, and the correspondence between metadata elements and metadata values needs to be clear from the information in the file itself. XML would seem to be the candidate format for such a purpose.

Level 0 – Printout of the metadata

Level 1 – Export the data to a standard XML format.

Level 2 – As with Level 1 but restricting elements to qualified and unqualified Dublin Core.

Level 3 – As with Level 1 but restricting elements to only unqualified Dublin Core.

Section II: Recommendations for a Preservation Plan for Hosted Assets

Preservation Recommendations

The Wisconsin Historical Society does not offer preservation of the archival copies of a Hosted Content Provider's digital assets. The Society does, however, promise that it will preserve the derivative copies of the assets a Hosted Content Provider uploads to the CONTENTdm server for disaster recovery purposes. While Hosted Content Providers must choose a preservation path within the boundaries of their own resources, the working group strongly recommends users consider the following combination as WHO "good practice":

Scanning

Scanning for WHO projects should be done at LEVEL 1 or higher.

Image Enhancement

Image enhancement for WHO projects should be done at LEVEL 1, which restricts image enhancement to cropping, rotation (de-skewing), levels or curves, and basic color correction.

File Format

WHO projects should be created using LEVEL 2 file format, which means that all archival files are saved in TIFF format using no or lossless compression, but avoid compression for color and grayscale images; bitonal images may use CCITT-Group IV compression.

Physical Format

WHO project archival files should be stored at least at LEVEL 1 and moved to LEVEL 4 when possible. Level 1 requires burning the images to CD with verifications made. LEVEL 2 requires CD burning with more complex verification. LEVEL 3 requires backing up preservation images to duplicate (or triplicate) magnetic tapes, with a provision for archival storage in multiple locations. LEVEL 4 will include the same activities as LEVEL 3 but will also include the storage of preservation images on a RAID device.

Metadata

Since WHO has embraced CONTENTdm, this means exporting the Dublin Core metadata into XML (CONTENTdm provides an XML export) to include standard and user-defined metadata elements. WHO metadata exports should be requested be requested by the content provider and procured periodically from the project host as collections are updated, and should be stored as LEVEL 3.

Section III: Additional Information for Migration & Refreshment of File Formats & Media

---

"[Digital files] reside on unstable media and must survive repeated processes or migration or perhaps emulation which have yet to be fully worked out, and which require long-term commitment to funding of perpetual maintenance."[5]

As recommended in Section I, migration and refreshment to new media and formats should be considered as part of a successful preservation program. The following information is meant to aid Wisconsin Heritage Online Content Providers in developing migration activities.

File Formats

Institutions engaging in digital preservation are encouraged to take an inventory of current formats in use and in what quantities the formats are used. The institution should then strive for normalization of formats, choosing one particular format for every type of digital object and ensure that all future files created are saved in these formats. Any migration from one format to another should be documented. Institutions should monitor factors that may contribute to the need for migration, such as software upgrades (although software generally has some backward compatibility, much software is only compatible 1-2 generations back), changes in the industry (wide adoption of a certain format/movement away from a certain format) and the availability of software that reads the current formats in use.

Since migration can lead to loss of data, some amount of risk assessment will have to be made when choosing to migrate. Ideally, software used for migration should have the ability to:

- Analyze the differences between the source file and the target format.
- Identify and report the degree of risk if a mismatch between formats occurs
- Accurately convert the source file to the target specifications

Emulation, the process of using current software and hardware to reproduce an older computing environment in order to access legacy file formats, should only be considered if migration is impossible.

Storage Media

As indicated in the Section I, CD-ROMs, magnetic tape, and RAID systems are recommended for storage. The effectiveness of DVDs and Solid State memory for digital preservation has not been established; therefore, these media are not currently recommended.

If using CD-ROMs, "gold/gold" CD-Rs (CD-Rs which use a gold reflective layer and phthalocyanine-based dyes) are shown to have the longest lifespan. Rewritable CDs should be avoided to prevent accidental erasure. Although CD-ROMs have been available since 1984, standards have changed throughout the years, so early CDs

---

[5]Janet Gertz, "Selection Guidelines for Preservation," *Joint RLG and NPO Preservation Conference: Guidelines for Digital Imaging.* http://worldcat.org:80/arcviewer/1/OCC/2007/09/28/0000073852/viewer/file69.html

may be completely obsolete. Although manufacturers claim a lifespan of anywhere from 70 to 200 years, it is advisable to check the physical condition of CD-ROMs after 5 to 10 years.

If using magnetic tape, Digital Linear Tape (DLT) and Linear Tape Open (LTO) are considered to be the most stable and longest-lasting forms of magnetic tape. They are believed to have a lifespan of approximately 30 years. When new generations of tape formats are offered, backward compatibility is usually only available for 1-2 generations back.

Proper storage of media can reduce the risk of physical deterioration and again, storage of one copy of project assets offsite (e.g., a safety deposit box) is recommended to reduce loss due to natural disaster or accidental destruction.

Recommendations for Proper Storage

- Maintain a constant temperature (68^{o} F is ideal)
- Maintain a consistent relative humidity as close to 40% RH as possible
- Avoid large or rapid fluctuations in temperature and humidity
- Avoid exposure to magnetic fields (for magnetic media)
- Avoid exposure to dust, fumes and cigarette smoke
- Prohibit food and drink in storage areas
- If possible, store media in closed metal cabinets that are electrically grounded
- Shelve media vertically
- Store media in their original cases
- Minimize exposure to light

Section IV: Glossary of Terms

Access: Files are maintained in a form that can be found by users over a network with appropriate retrieval engines. With appropriate output devices, users can view, print, listen to or otherwise access the files.

Archival file or copy: The full quality digital file from which lower quality **derivative files or copies** are made for delivery to a user on screen. For images, the archival file is often a TIFF file. Archival files, also known as archival masters or digital masters, are generally stored off-line and not mounted on a server. [6]

Bitonal: An image consisting only of a foreground color and a background color.

Born digital: A document that was created and exists only in a digital format.

Compression: Changing the resolution or size or eliminating data thought to be non-significant in order to make the file size smaller. See also **lossless** and **lossy**.

Derivative file or copy: The digital file that is presented to a user on screen. In most cases, these files are produced by reducing the size and quality of the **archival file or copy** through different means, including compression.[7]

Digital asset: Written, graphic or audio content in digital formats.

Digital preservation: Action taken to preserve the integrity, prevent deterioration and renew the usability and accessibility of items stored in digital electronic form and deemed worth maintaining for future generations. Digital preservation seeks to protect the original sources when they are deemed fragile, fulfill most, if not all, of the research and learning potential of the originals and, in some cases, transcend the original by enhancing with special lighting to draw out details, or higher resolution or incorporating searchable full text.

Emulator: Software that mimics old applications and file formats but which runs on current hardware. With a proper emulator, applications like Wordstar could run on today's computers.

File format: The format in which an electronic or digital asset is saved. Generally the extension of a file name will include the file format, and can be standard or **proprietary**. Examples are pdf, tif, jpg, gif, doc, xls, xml, htm, and the like.

Grayscale: A range of shades of gray in an image. Gray scales of scanners are determined by the number of grays, or values between black and white, that they can recognize and reproduce.

Image resolution: How much image detail an image can hold; a higher resolution means more image detail and larger file size.

[6] Modified from the Collaborative Digitization Program's glossary of terms
[7] Modified from the North Carolina "Exploring Cultural Heritage Online's" glossary of terms

Lossless: When files are compressed and then decompressed, a lossless compressed image is identical to the image before it was compressed, i.e., there is no data loss.

Lossy: When files are compressed and then decompressed, a lossy compressed image is not identical to the image before it was compressed and has less information. Some of the information deemed to be unimportant (e.g., color differences that are imperceptible to the human eye) is thrown away during the compression process. Future technologies may be able to distinguish these bits, so lossy compression may make certain uses of images unavailable in the future.

Metadata: Information about an item. This generally includes the name of the resource, who created it, when it was created and other important descriptive information. It may also include technical information such as the scanner used to create the digital version, decompression level, applications needed to view the work and files or databases that have more extensive descriptive metadata about this work (this is particularly important in the event that the digital file and its external metadata become separated). Extensive metadata is important to keeping an item accessible.

Migration: An approach that involves periodically moving files from one file-encoding format to another that is usable in a more modern computing environment. (An example would be moving a WordStar file to WordPerfect, then to Word 3.0, then to Word 5.0, then to Word 97.) Migration seeks to limit the problem of files encoded in a wide variety of file formats that have existed over time by gradually bringing all former formats into a limited number of contemporary formats.[8]

Pixel: The basic unit of the composition of an image on a television screen, computer monitor, or similar display. Generally, the smallest addressable unit on a display screen or bitmapped image.

Proprietary file format: A **file format** which is covered by a patent or copyright which is intended to give the license holder exclusive control of the technology.

RAID device: Redundant Array of Independent Disks. A disk subsystem that is used to increase performance or provide fault tolerance or both. RAID uses two or more ordinary hard disks and a RAID disk controller. In the past, RAID has also been implemented via software only.

Redundancy: Storing copies of records in more than one location or medium in order either to ensure a backup is available or for more efficient retrieval.

Refreshing: Periodically moving a file from one physical storage medium to another to avoid the physical decay or the obsolescence of that medium. Because physical storage devices (even CD-ROMs) decay, and because technological changes make older storage devices (such as 8" floppy drives) inaccessible to new computers, some ongoing form of refreshing is likely to be necessary for many years to come.[9]

Resolution: See **image resolution**

---

[8] Howard Besser, "Digital Longevity," *Handbook for Digital Projects: A Management Tool for Preservation and Access* (Andover, MA: Northeast Document Conservation Center), 2000. http://www.nedcc.org/resources/digitalhandbook/dman.pdf paragraph 26

[9] Besser.

SGML: Standard Generalized Markup Language. A system standard for defining the format in a text document. SGML uses a separate file that defines the format codes, or tags, embedded within it. Since it describes its own formatting, it can be used in a wide variety of applications. HTML is an SGML document that uses a fixed set of tags.

XML: Extensible Markup Language. A flexible way to create common information formats and share both the format and the data in a consistent way on the web or a network. XML is a simplified version of SGML.