Best Practice
Digital Resource Maintenance and Preservation Strategies

Responsible Office: Yale University Library	**Effective Date**: 4/9/2007
Responsible Official: Digital Preservation Committee	**Revision Date**: 4/9/2007

Scope

This best practice defines the set of interdependent strategies that may be used to maintaining access to digital resources over the long-term. This includes strategies for maintenance in the short to medium term based on safeguarding storage media, content and documentation, as well as strategies for long-term preservation to address the issues of software and hardware obsolescence. Many different digital preservation strategies have been proposed and developed by many different practitioners and researchers around the world. No one strategy is appropriate for every type of digital resource, every situation, or every organization.

Reason for this Best Practice

While other kinds of library resources may survive periods of benign neglect, digital resources will not. Digital resources require well planned, well managed, and sustained strategies over their entire existence (or as long as Yale wishes to be responsible for their existence). Because not all digital resources in the Yale University Library necessitate long-term preservation, not all digital resources require such preservation strategies. However, all digital resources require maintenance strategies in order to keep them accessible and protect them from imminent threats like media, hardware and software failure, operator error, natural disaster, external or internal attack, and economic or organizational failure.[i] This document presents such a list of maintenance and preservation strategies; they are largely drawn from the UNESCO *Guidelines for the Preservation of Digital Heritage*.[ii]

Defined Digital Collections

Before any strategies can be undertaken to maintain and preserve digital resources, it is necessary to clearly define what digital resources <u>are</u> and what digital resources <u>are not</u> part of the organization's collections. This is essential as many digital preservation discussions are drawn off track by issues relating to digital resources that are not the responsibility of the organization to preserve over the long-term. Yale and YUL do not yet have properly defined digital collections. Existing descriptive and collection management systems may not be sufficient to define digital collections. Ad hoc efforts, like the Provost's digital landscape survey have been necessary to gain insights into the scope of collections.

A. <u>Maintenance Strategies</u>

A maintenance strategy prescribes actions to be taken to eliminate problems with the storage of digital resources during the earlier stages of their existence. While much attention is paid to the development of complex long-term preservation strategies, such strategies are inapplicable if the digital resources are not properly stored, maintained and protected. These strategies are the minimum necessary requirements for storing, protecting, and maintaining the accessibility of authentic copies of digital resources through the earlier stages of their existence. If any of these strategies are not implemented as part of a well planned, well managed, and sustained program, the digital resources will be at risk. There are eight primary maintenance strategies. All of them are necessary to ensure reliable and accurate storage of the digital resources' components so that they will exist long enough for preservation strategies (as listed in section B) to become effective. The maintenance strategies are:

A1. <u>Clear allocation of responsibilities</u>. A person or office must be given unambiguous responsibility for managing digital resource storage and protection of all digital collections. This is a technical responsibility that requires a specific skill set, dedicated resources, and an appropriate plan. This strategy can be undertaken by hiring a competent staff member devoted to this task or by assigning to it a specific portion of the time of existing staff or of an existing office. Except for the very smallest of digital collections, it is recommended that a staff member (or group of staff members) clearly be identified as the responsible party. The larger the institution and the greater the size of the digital collections; the more staff that will be necessary to manage the storage and protection of the organization's digital resources.

Existing YUL Practice: There is no single person or unit responsible for ensuring the storage and protection of all digital collections. The Library Integrated Access Council appointed the digital preservation committee charged to ensure that digital information of continuing value remain persistent, accessible and usable. ILTS has general purview over computer and network security and manages the Rescue Repository. The recently formed Digital Repository Service establishes a proper framework for future development I this area.

Recommendations for Future Practice: Somebody needs to have a clearly allocated responsibility in every unit for every digital collection deemed worthy of preservation. There will need to be a better connection between selectors generally responsible for collections and the assignment of technicians responsible for storage and protection.

A2. <u>Provision of the appropriate technical infrastructure</u> and of the auxiliary staff necessary to manage storage. Technical infrastructure includes all of the physical and administrative resources, which enable the storage management process (buildings, computer hardware, computer networks, and the auxiliary staff necessary to maintain the same).

Existing YUL Practice: The Library is beginning to develop the framework for the necessary infrastructure. Administrative analyses of technology units; hiring of competent professionals; encouragement of professional development; purchase of hardware and development of software for storage; and development toward a digital repository utilizing Vital/Fedora are all positive developments, but much more must be done. Departmental silos restrict much of the infrastructure. Some units have invested a great deal into infrastructure development and hiring of staff with requisite technical skills. Yet there is little library-wide effort to examine technical skills across all new hires and common understanding of how to pool resources to build the large-scale repository infrastructure given existing budgets.

Subsections:

- System Infrastructure Best Practice
- Operating System and Software Best Practice

A3. <u>Establishment and implementation of a plan for system maintenance, support and replacement</u>. This includes the maintenance, updating, and/or replacement of hardware and software

Existing YUL Practice: Such maintenance is implied every 3-5 years, but there is no written policy and it is not necessarily uniformly applied across the board, especially when storage is not managed in the Rescue Repository.

A4. Establishment and implementation of a plan for regular transfer of records to new storage media on a regular basis. The copying of digital components from one storage medium to another in order to avoid the impact of media decay. Such transfer should be undertaken in a systematic manner.

Existing YUL Practice: There is an informal recognition of a schedule for media refreshing for collections and applications managed by ILTS. For example, the purchase of Orbis included plans for media refreshing. However, there is no systematic program implemented across all digital collections.

A5. Adherence to appropriate storage and handling conditions for storage media. Adhering to appropriate conditions can dramatically reduce the rate of media decay. For instance, excessive heat, humidity, and dust endanger storage media. Keeping storage environments according to established standards is a necessary strategy to ensure the protection of the digital components of records.

Existing YUL Practice: The situation here is similar to that for A5. This is done for mission critical systems by ILTS without an explicit set of procedures.

A6. Ensuring redundancy and regular backup. The duplication of digital entities and the storage of the resulting multiple copies on different physical media and/or in different physical locations protects them against media failure that may result in one or all physical media in a particular storage site location.

Existing YUL Practice: This is done for major applications. There are not written guidelines. Nobody is checking on adherence.

A7. Establishment of system security. Controls should be implemented to ensure that digital components of records are exposed only to controlled, authorized users and/or processes.

Existing YUL Practice: At the network level, there is tremendous investment by ILTS and ITS devoted to system security. Yale University Policy 1610 establishes IT security requirements for faculty, students, staff, and other individuals who use computing or communications systems during the course of their work at Yale. This includes systems used on-campus as well as from remote locations, such as home, hotels and other off-campus locations.

A8. Disaster planning. The strategies listed above are designed to minimize accidental loss of the digital components of records and maximize media longevity, but even with perfect storage conditions and excellent handling protocols, disasters may still happen. A disaster recovery plan should contain detailed procedures for restoring a damaged system and for guiding in the effective recovery of recordkeeping systems following a disaster.

Existing YUL Practice: ITS is dealing with disaster planning at the highest levels.

B. Preservation Strategies. In addition to the maintenance strategies, every preserver is responsible for establishing a trusted preservation system and for establishing and implementing preservation strategies for all digital resources to be kept for the long term. Preservation strategies prescribe actions to be taken to eliminate problems of storage over the long term. All possible preservation strategies are listed as below in order to describe the range of possible preservation choices. While they are listed individually, it is most likely that in practice a combination of two or more different preservation strategies most likely will be necessary for each collection of digital resources.

B1. Use of standards: This strategy involves the use of widely available, supported or agreed standards and/or file formats, for which there is an increased likelihood of stability and longer-term support. Such standards or formats may either be formally agreed or may be de facto standard formats that have been widely adopted by industry. Compliance to standards may also either simplify the application or maximize the effectiveness of later preservation strategies. This strategy can be related to B4 (Restricting the range of formats to be managed). A particular refinement of the standards approach is referred to as durable encoding, which recommends encoding data to confirm with well-known data processing standards down to the level of encoding bits as ASCII or Unicode UTF-8, and objects as XML.

Advantages: standard formats are more likely to have a range of tools to work with; some of the effort necessary for preservation can be off-loaded onto the producer.

Disadvantages: The investment necessary to deal with each standard (if other institutions don't step forward to deal with preservation of YUL chosen standard); existing standard formats may not be capable of representing all of the elements of the formats which they are substituting for; and acceptable standard formats may not yet exist; preserver may not be able to influence the producer of digital resources; may give false sense that preservation efforts have been completed; standard file formats are not always well chosen; some YUL staff must be knowledgeable about and monitor standards developments; YUL selectors must be able to choose standards, both with regard to the effect of any transformation on the essential characteristics of digital resources and the expected longevity of tools to work with these standards.

Existing YUL Usage: DL and MADID restrict file formats to defacto standards

Recommendations for future use: common metadata standards; file format standards for preservation mater and use copies of file types (audio, video, still image, text, etc.);

B2. Data extraction and structuring (self-describing formats, persistent object preservation, tagging): This strategy involves analyzing and tagging data so that the functions, relationships and structure of specific elements can be described. The re-presentation of content can be liberated from specific software applications and be achieved using different applications as technology changes.

Advantages: infrastructure independence; simplifies the transfer of digital resources between generations of technology

Disadvantages: expensive and time consuming effort in the beginning (pay for preservation up front); requires extensive development of tools and methods for analysis and processing; exact presentation characteristics may be difficult to reproduce

Existing YUL Usage: No existing usage.

Recommendations for future use: This strategy is most promising for social science data sets.

B3. Encapsulation: This strategy involves binding together data and the means of providing access to that data, normally in a 'wrapper' that describes what it is in a way that can be understood by a wide range of technologies (such as an XML document). Encapsulation often bundles metadata describing or linking to the correct tools. This is the way that the OAIS Reference Model describes information packages for preservation.

Advantages:

Disadvantages:

Existing YUL Usage: No existing usage.

Recommendations for future use: This strategy will require further research and development before it can be recommended for use in YUL.

B4. Restricting the range of formats to be managed (normalization): Only storing data in a limited range of formats. This can be achieved either by only accepting material already in those formats, or by converting material from other formats before storage. This is a refinement of B1.

Advantages: By limiting the total number of different formats that must be preserved, the organization can limit the technical infrastructure necessary for preservation.

Disadvantages: A limited number of file formats may not be capable of representing the desired look, feel, and functionality of all digital resources selected for preservation.

Existing YUL Usage: Digital surrogate still images are currently maintained in a small number of file formats.

Recommendations for future use: Continue to utilize for access copies of images and also employ for access copies of video files.

B5. Technology preservation: This strategy involves keeping and maintaining the original software and hardware with which electronic records were presented.

Advantages: presenting digital resources through their intended hardware and software ensures the full range of intended elements and functions are presented

Disadvantages: as time goes by parts and maintenance expertise become increasingly scarce; requires wide-ranging YUL expertise to use and maintain the technologies on site; requires management of licenses and agreements; knowledge tends to be held by a single staff member; copyright issues for software

Existing YUL Usage: MSSA and BRBL maintain old disk drives that will accommodate disks of a size that are no longer accommodated by current computer equipment; MSSA is also maintaining superseded software for use with legacy materials and old operating systems to support software that does not work on current platforms.

Recommendations for future use: when preserver must be able to deal with any technology the producer is working with (University Archives); for complex artistic or scientific digital resources where the technology is integral to the representation of the digital resource; if the technology holds some value as an artifact (Beinecke).

B6. Reliance on backward compatibility: This strategy relies on the ability of some software to interpret and present digital components of records created with previous versions of the same software. In this strategy, the presentation is limited to temporary viewing, whereas migration permanently converts records into a format. This is often used in conjunction with B1.

Advantages: This strategy requires only minimal effort as long as software applications are wisely chosen for longevity.

Disadvantages: The risk with this strategy is that those responsible for software development may not hold backward compatibility as essential. The library may be powerless to prevent dramatic alterations to software applications that may prevent full backward compatibility.

Existing YUL Usage: The most common example of this strategy is the use of Microsoft Word word processor with .doc format files. All older .doc files can be read in all newer versions of Word (that is, until the Windows Vista version of Word).

Recommendations for future use: Because of its ease and limited expense, this is a effective strategy as long as there is proper recordkeeping of file formats in use and their software application dependencies. This will enable more effective decision-making regarding alternate strategies to pursue when software applications change.

B7. Migration: Migration involves transferring digital materials from one hardware or software generation to another. As distinct from refreshing, which maintains the data stream by transferring it from one carrier to another, migration entails transforming the logical form of a digital object, so that the conceptual object can be rendered or presented by new hardware or software. The most commonly proposed migration method involves permanently transforming one logical format into another in line with technological change, so that all migrated objects can be presented with prevailing technology. It is also possible to propose a 'migration on demand' or 'migration at the point of access' model. This approach is discussed under B9 ('Viewers') below.

Advantages: Migration addresses issues of obsolescence and media decay at once by altering the encoding and format of the data.

Disadvantages: Migration can be expensive and time consuming to undertake. Also, "in practice, migration is prone to generating obvious and subtle errors. An obvious error occurs when the set of structural elements in the source format does not fully match the structural elements of the target format. For instance, in a spreadsheet file a structural element defines a cell containing a numeric value. If a comparable element is missing from the format specifications of the target format, data will be lost. A subtle error might occur if the data themselves do not convert properly. Floating point numbers (numbers with fractions) are found in many numeric files. Some formats might allow a floating-point number of 16 digits (e.g., 26.00126700998l9070) while others might allow only 8 digits (e.g., 26.00126701). For some applications, such as vector calculations in geographic information system (GIS) programs, small but significant errors could creep into calculations. In other situations, migration might preserve the content of the file but lose the internal relationships or context of the information. For example, a spreadsheet file migrated to ASCII may save the current values of all the cells but lose any formulas embedded within the cells that are used to create those values." [iii]

Existing YUL Usage: The data in the OPAC was migrated to the current version several years ago. Many different databases are periodically migrated from an obsolete or unsupported data format to a new format. For example, Manuscripts and Archives is currently migrating collection management data from Paradox, Access, Excel, and MARC to MySQL to work with the Archivists Toolkit application.

Recommendations for future use:

B8. Software re-engineering: Transforming software as technologies change, similar to the transformation of data formats. This may include anything from the re-compiling of source code for a new platform to the re-coding of the software from scratch in another programming language.

Advantages:

Disadvantages:

Existing YUL Usage:

Recommendations for future use:

B9. Viewers and migration at the point of access: The use of software tools or transformation methods that provide accessibility at the time of access, using the original data stream.

Advantages:

Disadvantages:

Existing YUL Usage: MSSA and BRBL utilize Quick View Plus to provide access to a number of digital resource file formats.

Recommendations for future use:

B10. Emulation: Using software that makes one technology behave like another (making future technologies behave like the original environment of a preserved digital object, so that the original object could be presented in its original form from the original, or migrated data streams).

Advantages: Users experience the full look, feel, and functionality of the original environment.

Disadvantages: Very expensive and difficult process.

Existing YUL Usage: No existing usage.

Recommendations for future use:

B11. Non-digital approaches: A strategy to 'print out' the digital resources and/or their related information onto relatively stable analogue media, such as paper, microfilm, film, audio and video tape, shifting the preservation burden to an analogue copy in place of the digital object or, in the case of digitized resources, preserving the analogue original rather than preserving the digital surrogate copy.

Advantages: provides a simpler preservation alternative, as analogue materials may be preserved for the long-term using traditional preservation methods; digital resources are captured in human-readable form and are removed from the threat of technological obsolescence and the pressure of ongoing digital preservation cycles

Disadvantages: lose advantages afforded by digital technology such as convenience of use, search and navigation features, or storage efficiency; long-term stability of analogue carriers may depend on expensive storage environments

Existing YUL Usage: MSSA is exerting very little preservation effort on its surrogate digital copies, depending instead on the analogue originals

Recommendations for future use: only suitable for objects that do not require the functions of digital technology to achieve their purpose

B12. Data restoration (archaeology): Recovering data as bits from physical media followed by steps to restore the intelligibility of the recovered data. It is most often employed in recovery of data from failed, damaged or degraded media, but methods to restore intelligibility have been used to rescue documents in obsolete formats. This would be the equivalent of the deciphering ruined papyri after 2000 years.

Advantages: This strategy has the benefit of procrastination. If an organization is planning to rely entirely on data archaeology, most work can be put off until the future. If there is very little budget for digital preservation, an organization may need to hope that future data restoration can be accomplished successfully. In some cases this may be the only possible strategy to attempt if the digital resources have been neglected.

Disadvantages: This is a high-risk strategy. The gamble is that technological development may allow some better way to reconstruct the data. One is also gambling that the decay of the media has not progressed so far as to make the restoration impossible. Also, this strategy is very expensive and labor intensive. It would be impossible to apply this strategy across all types of digital resources. It could only be attempted for only the most extraordinarily important digital resources.

Existing YUL Usage: No existing usage.

Recommendations for future use: Only special collections units acquiring very important collections that have already become obsolete, corrupt, or have experienced media decay should utilize this strategy.

---

[i] David S. H. Rosenthal, Thomas Robertson, Tom Lipkis, Vicky Reich, Seth Morabito, “Requirements for Digital Preservation Systems: A Bottom-Up Approach,” *D-Lib Magazine*, 11:11(November 2005) <http://www.dlib.org/dlib/november05/rosenthal/11rosenthal.html>.
[ii] *Guidelines of the Preservation of Digital Heritage*, CI-2003/WS/3, United Nations Educational, Scientific, and Cultural Organization (UNESCO), Information Science Division, Prepared by the National Library of Australia, March 2003 <http://unesdoc.unesco.org/images/0013/001300/130071e.pdf>.
[iii] Gregory W. Lawrence, William R. Kehoe, Oya Y. Rieger, William H. Walters, and Anne R. Kenney, *Risk Management of Digital Information: A File Format Investigation*, Council on Library and Information Resources (pulication 93), June 2000 <http://www.clir.org/pubs/reports/pub93/contents.html>.