Preservation and Transfer Media and Formats:
Making Responsible Choices

Jim Suderman
Presented to the SARBICA Conference, Hanoi, 5 May 2004

The choice of preservation and transfer media and formats is a practical topic and as such any discussion of it must take place within the specific preserver's context – particularly in relation to the resources available to the preserver. It is not particularly governed by theoretical principles. The first part of this paper will review media chosen for transfer and preservation of electronic records at leading institutions in Canada, the United States and the United Kingdom.[1] The second part will review transfer and preservation formats. Essentially these sections are simply lists. The final section will briefly outline some considerations or guidelines for choosing preservation media and formats. This is where the "Making Responsible Choices" element in the session title will be addressed.

Transfer and Preservation Media

File Transfer Protocol (FTP) – Currently only "unclassified, uncompressed, unrestricted, and unencrypted files" are accepted via FTP at NARA.[2] Library and Archives Canada is considering a recommendation to use FTP and Web Distributed Authoring and Versioning (WebDAV) Protocol as online transfer mechanisms.[3]

Tape media (transfer or storage)[4]

- o Digital Linear Tape (DLT IV) cartridge tape (non-compressed);

- o Linear Tape – Open (LTO) cartridge tape

- o 8mm tape[5]

- o Open reel 9-track tape

- o 3480 class tape cartridges

- o Compact Disc – Recordable (CD-R)

- o 'Spinning disk' (online or near line storage)

---

[1] Please note that this information is believed by the author to be accurate at the time of the presentation, but should not be relied upon for decision-making. Institutions considering selecting preservation or transfer media should make direct enquiries to other electronic records repositories.

[2] "Expanding Transfer Options for Electronic Records" at NARA website, accessible at http://www.archives.gov/records_management/initiatives/transfer_regulation.html, checked 18 March 2004).

[3] Mike Swan. "Library and Archives of Canada: Electronic File Transfer Analysis" (2003), pp. 4-5. This report also examines security measures and assesses applications for user interface.

[4] This listing includes all media used for storage or transfer or electronic records at any one of the United States National Archives, the National Archives of the United Kingdom, and Library and Archives Canada.

[5] Library and Archives Canada initially invested in the STORLORD optical platter as a preservation media. This media is now obsolete and holdings are being transferred to 8mm tape. This process may have been completed by the time of this presentation.

This latter storage media is currently in use in some institutions, including the National Archives of Australia, I believe. In my institution disk storage has succeeded tape media and was brought about by a realization that we did not have the needed skills to efficiently and reliably store records on tape media. Combined with this was an upgrade in our technology base, from Windows NT to Windows 2000. The preservation tapes had been made in the NT environment and were inaccessible in the 2000 environment, necessitating an immediate resolution.

Transfer and Preservation Formats

In 2002 the National Archives in the United Kingdom initiated a remarkable new project named PRONOM. The PRONOM Project gathers and makes available technical information about file format structures and supporting software products. This is seen as a prerequisite for any digital preservation regime. This preservation approach represents a significantly different approach to preservation from that proposed by the National Archives of Australia's green paper (related in an earlier session). It is similar in that it represents a significant and on-going institutional commitment to retain particular knowledge and specifications as part of the preservation program. Part of this commitment is establishing and maintaining contact with major software developers for software product information. Both approaches appear to embrace the principle that most of the investment in preservation should be made at the outset, rather than rely on future developments in technology to address preservation concerns. The plans for the coming years would see PRONOM enhanced to include "specific tools to support digital preservation activities."[6]

The following listing of transfer formats is from Library and Archives Canada.[7] Like PRONOM, this initiative represents a significant institutional investment as the formats identified in this guideline are subject to the rapid transitions of technological developments and changes in the market. For this reason the guideline is scheduled for annual review with updates as necessary. The reasons for the selection of these formats are as applicable from a preservation standpoint as they are from the standpoint of facilitating a transfer of records.

While Library and Archives Canada has the technological capacity to handle all the file formats identified, the formats are either 'recommended' or 'acceptable.' Recommended formats are those that support transfers both of records to the Archives, and between governmental agencies, i.e., while records are still active. Acceptable formats are those that should only be used to transfer records to the Archives. In its

---

[6] See "Background" at the PRONOM website at http://www.records.pro.gov.uk/pronom/background.htm (checked 18 March 2004).

[7] Electronic Records Development Division, Library and Archives Canada. "Guidelines for Computer File Types, Interchange Formats and Information Standards" (Version 1.0, February 2004). The [UK] National Archives "does not specify or endorse the use of any particular file formats." See Adrian Brown, *Digital Preservation Guidance Note 1,* "Selecting File Formats for Long-Term Preservation," p. 4, available at http://www.pro.gov.uk/about/preservation/digital/guidance/default.htm, checked 22 March 2004.

selection of file formats, the requirements for record quality, stability, longevity, and industry acceptance were balanced. Note that the distinction between recommended and acceptable formats is based purely on technological, as opposed to business process, reasons.[8]

Character Sets:

- o American Standard Code for Information Interchange (ASCII) [ISO/IEC 8859-1:1998 (Latin-1)]

- o Extended Binary Coded Decimal Interchange Code (EBCDIC)

- o Unicode Version 3.0 UTF-8 [ISO/IEC 10646-1:2000]

At the present time NARA accepts only ASCII and EBCDIC character sets, although consideration of other formats is underway.[9]

- o Digital Audio

  - o Audio Interchange File Format (AIFF)

  - o WAVE (WAV)

  - o MPEG-1: Layer 3 (MP3)

  - o Musical Instrument Digital Interface (MIDI)

  - o Real Audio (RM/RA)

- o Digital images

  - o International Telecommunication Union-Telecommunication Standardization Sector (ITU-T) T.4 and T.6

  - o Portable Network Graphics (PNG)

  - o Tagged Image File Format (TIFF)

  - o Graphics Interchange Format (GIF)

  - o Joint Photographic Experts Group (JPEG) [ISO/IEC 10918-1:1994]

  - o JPEG File Interchange Format (JFIF)

---

[8] From the Introduction to "Guidelines for Computer File Types, Interchange Formats and Informations Standards", p. 7. For enquiries concerning the specific breakdown of which formats are recommended, please contact the Electronic Records Development Division, Library and Archives Canada, 344 Wellington Street, Ottawa, Canada, K1A 0N3 (email: imgi@archives.ca).
[9] See NARA Regulations in the Code of Federal Regulations, § 1228.270 Electronic records at http://www.archives.gov/about_us/regulations/part_1228_l.html, checked 22 March 2004.

Jim Suderman

NARA accepts JPEG/JFIF and TIFF formatted images, subject to certain image quality specifications.[10]

- o Digital video

    - o Moving Pictures Expert Group (MPEG-2)

    - o Audio Video Interleave (AVI)

    - o MPEG-4

    - o Quicktime (MOV)

    - o Real Networks' RealVideo (RM)

- o Documents – Textual

    - o Extensible Markup Language (XML)

    - o Extensible HyperText Markup Language (XHTML)

    - o HyperText Markup Language (HTML)

    - o Standard Generalized Markup Language (SGML) [ISO/IEC 8879:1986]

    - o Text Files (.txt)

    - o Microsoft Word Document Format (.doc)

    - o Portable Document Format (PDF)

    - o WordPerfect Document Format (.wpd)

With increasing frequency textual documents exist in scanned image formats. NARA will accept scanned images of textual documents in for TIFF, GIF, BIIF (Basic Image Interchange Format), and PNG formats, subject to certain quality specifications.[11] Similarly, NARA will accept records created and converted to PDF format, subject to certain requirements, e.g., removal or deactivation of security settings.[12]

- o Email

    - o Multipurpose Internet Mail Extensions (MIME)

---

[10] See "Expanding Acceptable Transfer Requirements: Transfer Instructions for Permanent Electronic Records - Digital Photographic Records" at
http://www.archives.gov/records_management/initiatives/digital_photo_records.html, (checked 22 March 2004).
[11] "Expanding Acceptable Transfer Requirements: Transfer Instructions for Existing Permanent Electronic Records - Scanned Images of Textual Records" at
http://www.archives.gov/records_management/initiatives/scanned_textual.html, (checked 22 March 2004).
[12] "Expanding Acceptable Transfer Requirements: Transfer Instructions for Permanent Electronic Records - Records In Portable Document Format (PDF)" at http://www.archives.gov/records_management/initiatives/pdf_records.html, (checked 22 March 2004).

Jim Suderman

NARA will accept email in standard markup language (preferred) or native formats, both subject to certain specifications. PDF format is explicitly identified as an unacceptable format for email, although that format may be acceptable for attachments.[13]

- Geospatial Data

  - Digital Line Graphs – Level 3 (DLG-3)

  - Environmental Systems Research Institute (ESRI) Export Format – (E00)

  - Environmental Systems Research Institute (ESRI) Shape Format – (SHP)

  - GeoTIFF

  - Geography Markup Language (GML), Version 3

  - International Hydrographic Organization (IHO) S-57, Edition 3.1

  - TC 211 ISO 191xx Standards for Geographic Information

  - Spatial Data Transfer Standard (SDTS)

  - Canadian Council on Geomatics Interchange Format (CCOGIF)

  - CARIS ASCII

  - CEOS Superstructure Format

  - Digital Elevation Model (DEM)

  - GeoVRML (Virtual Reality Modelling Language)

- Structured Data – Databases and Spreadsheets

  - Flat File

  - dBase Format (DBF)

- Technical Drawings

  - Drawing Interchange File Format (DXF)

Making Responsible Choices

Ultimately the choice of preservation formats and media must meet the responsibilities of the archival program: the secure preservation and reliable reproduction of electronic records. Those responsibilities are met through choices that are based on the technologies that are accessible to us and that our institutions can afford, through knowledge of the effect of these technologies on the records and access to them. The

---

[13] "Expanding Acceptable Transfer Requirements: Transfer Instructions for Existing Email Messages with Attachments" at http://www.archives.gov/records_management/initiatives/email_attachments.html, (checked 22 March 2004).

Jim Suderman

Australian green paper cited previously suggests that "most of the preservation effort needs to be invested at the beginning, not in continual emulator maintenance or data conversion."[14] I understand this to mean that part of the preservation program's responsibility is not only to stabilize electronic records for preservation, but also stabilize the ongoing cost of maintaining them within the archival repository.

The rate of technological change means that any selection of media or format for preservation is a temporary one. Because there is no particular archival emphasis on the medium, making a responsible choice here is based primarily on factors affecting the institution. Because the archival endeavour is not just about preserving records, perhaps the primary archival consideration to take into account when selecting a preservation medium is its impact on providing access to the records.

The National Archives in the United Kingdom identifies server-based hard disk storage as "the most effective and secure storage regime for electronic records, provided it is well managed and includes an effective backup strategy."[15] It is acknowledged that such storage may not always be possible and identifies six criteria for assessing removable storage media. These are:

- o a proven lifespan of at least ten years;

- o a storage capacity appropriate for the amount of data to be stored;

- o high quality of error-detection and correction;

- o a mature technology with a well-established market presence;

- o a good return on investment (including the price per MB and total costs for purchasing and maintaining the necessary hardware and software;

- o be durable both in terms of the protective casings or coatings and, for magnetic media, with a high coercivity value.[16]

To these helpful criteria might be added

- o costs for new training for staff responsible for managing the storage media;

- o costs for storage space, e.g., converting shelving to accommodate a different media form factor;

- o the possibility that current work processes may have to be changed – note that new media may affect work processes for both preservation and the provision of access to the records;

---

[14] Helen Heslop, Simon Davis, Andrew Wilson. "An Approach to the Preservation of Digital Records" (National Archives of Australia Green Paper, 2002), p. 13. Available at http://www.naa.gov.au/recordkeeping/er/digital_preservation/Green_Paper.pdf, checked 22 March 2004.

[15] Adrian Brown. "Selecting Storage Media for Long-Term Preservation" The National Archives. *Digital Preservation Guidance Note: 2* (2003), p. 4. (Available at http://www.pro.gov.uk/about/preservation/digital/guidance/default.htm, checked 22 March 2004.)

[16] Recommended to be in excess of a thousand Oersteds.

The InterPARES baseline requirement that preservation of authentic electronic records requires knowing how records will be affected when they are reproduced makes selection of formats somewhat more challenging. Formats that are otherwise satisfactory may not reproduce the record satisfactorily. Satisfactory preservation formats may also complicate provision of access to records. For example, transfers of email from many different email accounts. It may be simplest to transfer all the email in a database format. If the entire email database was acquired, it may be difficult to limit access to individual email records, or the emails of just one user. If individual files reflecting the emails of each individual user are acquired, it may be difficult to access the entire body of emails. The choice of preservation format will affect how access to the records can be provided. Because this sort of assessment may be intensive, changing the format of records should be done as infrequently as possible, and the longevity of the format chosen should be an important consideration in this process.[17]

The choice of preservation media is dominated by the market, in terms of what media are available, and cost, i.e., the preserver can afford. These factors are also relevant when choosing preservation formats. In this latter case, however, the impact of the format on the ability to reliably reproduce authentic records is paramount. For these reasons both of these considerations are relevant in terms of the management of the archival appraisal and preservation business processes.

---

[17] "To lessen the risk to the integrity of the records, the preservation program will minimise the number of preservation treatments applied to each digital records." Heslop, Davis, Wilson. "An Approach", p. 16.

Jim Suderman