Transfer Policies of Electronic Records

Jim Suderman
Presented to the SARBICA Conference, Hanoi, 3 May 2004

The Preservation Model from InterPARES 1 identifies five factors affecting transfers of electronic records from a creator or user to a preserver.[1] These factors are the:

- the accessioning policy;
- the preserver's technological infrastructure (including available skill sets);
- a preservation strategy for a particular body of records;
- a targeted preservation method (i.e., how the strategy will be executed); and
- the actual digital components and accompanying information of the transferred records.

An accessioning policy is defined as the preserver's policy for accepting responsibility for transferred records. This policy identifies standards and specifications for transferred records. Thus it might address considerations such as

- approved transfer file formats (e.g., database records to be transferred in comma separated value format);
- what should be done when records are missing, or incomplete; or documentation is missing or unclear;
- a transition date setting out when responsibilities for providing access to the records move from the creator to the preserver.

With the electronic records program in its infancy, my institution's accessioning policy is primarily focussed around testing transfers in advance. Because no facility has been developed for an on-line transfer of records, electronic records are transferred on digital media such as tape or optical disc. Writing data to tape is usually done using backup software, which is highly proprietary, and so prior to the actual transfer, the media and format are negotiated. The agreed configuration is then tested and, if successful, is used as the basis for the actual transfer of records. At the same time, the chosen file formats are confirmed. A listing of acceptable file formats has not been finalized. Most office automation files are accepted in their native format, i.e., in the same format as that in which they were created or used. The preferred file format for images is tagged image file format (.tiff), but joint photographic experts group (.jpeg) and graphics interchange format (.gif) may also be accepted. The database format (.dbf) is the preferred format for databases. The comma separated value (.csv) format may be acceptable, especially for spreadsheet applications. These formats are accepted by a number of archives in Canada.

[1] The five factors refer to the arrows pointing to the "Bring in electronic records" (A2) activity identified in the "Preserve Electronic Records" model. InterPARES Preservation Task Force *Model diagrams*, available at http://www.interpares.org/book/interpares_book_n_app05ii.pdf, checked 29 March 2004. Definitions of the factors are from the Preservation Task Force *Arrow definitions*, available at http://www.interpares.org/book/interpares_book_n_app05iv.pdf, checked 29 March 2004.

Following the actual transfer, additional checks are undertaken beginning with a basic confirmation that the received records are, in fact, the records that were expected. Whenever possible we have also confirmed the total number of records. In one case there was a discrepancy of roughly 10,000 files from what we were expecting. We communicated this discrepancy to the transferring body and resolved the matter. This is important to confirm that we have received everything we need to have received to hold a complete record.

In addition to these technological considerations, there is usually a protocol established to govern requests for access to the records. In Ontario these are governed by the *Freedom of Information and Protection of Privacy* law. This protocol would also govern legal document discovery requests. Normally requests for information or records received up to the transfer date remain with the originator of the records. Following the date of the successful transfer of the records all requests for records become the responsibility of the Archives.

The technological infrastructure refers simply to the common hardware and software environment for executing preservation processes. My current technological infrastructure is currently limited to a PC-based environment, with Windows 2000 as the operating system. This environment will have to be upgraded in the not too distant future as the volume of holdings grows and as the need to provide access to the system increases.

A preservation strategy is defined as a "coherent and comprehensive approach for preserving a body of records selected for preservation…" It consists of objectives for

- maintaining components of electronic records and related metadata and information over time, expressed in terms of one or more action plans for different classes of digital components, indicating times or conditions for actions to be taken.
- reproducing the records in authentic form;
- criteria for evaluating execution of the strategy.

The reference to "components" here is a reflection of the conclusions of the Preservation Task Force:

> An electronic record is stored as one or more digital components. Digital components have no necessary relation to the elements of documentary form recognized in diplomatic analysis of records. Rather, they are determined technologically by the way the bits are stored and by the methods (software) that must be applied to reproduce the record. Reproducing an electronic record entails (1) reconstituting it, that is, reassembling its digital components if it has more than one, or extracting any digital component stored in a physical file that contains more than one such component; and (2) presenting it in proper form.[2]

[2] InterPARES Preservation Task Force Report, pp. 5-6, available at http://www.interpares.org/book/interpares_book_f_part3.pdf, checked 29 March 2004..

A familiar example of this is a web page, including an image or picture, viewed through a browser. The web page doesn't normally include the image file or component, but instead includes text that indicates how the image will appear on the page. The image itself is stored as a separate component from the web page. And so there are at least two components that must be assembled to reconstitute the page. Note that reconstituting the page in this way also requires a browser application. It is possible to open the HTML file in a text-processing application and the image file in an image-processing application, but only an application with browser like capabilities will be able to reconstitute the page in the way it was intended to be presented. The preservation strategy for the HTML file will differ in some details from that of the accompanying image component.

The targeted preservation method refers to the software used to implement a preservation strategy. This requires determining which types of records or classes of digital components the method will apply to, and the conditions of applying it.

The software used to implement a preservation strategy may be closely linked to an acceptable file format for transfer. For example, comma separated value (.csv) format files can be easily accessed through current spreadsheet applications such as Microsoft Excel. In the current environment, this may be a reliable way to reproduce the record for access. In this instance the preservation method does not require any kind of file format conversion although the reproduction of the record may be affected. I have recently begun experimenting with the Open Office suite of applications,[3] converting Microsoft Word, Excel, and Powerpoint formatted files to the Open Office xml-based format. An attractive feature about Open Office is that it provides a means to undertake batch conversions. However, any decision to commit to this kind of preservation method for all records would need extensive testing to ensure that it is a reliable method for preserving documents written in French, or with unusual formatting, etc.

Finally there are the actual digital components, the accompanying information related to preservation and reproduction of the related record, and transmittal information. The digital components must be transferred, of course. Documentation is likely to vary with the records being transferred. Transfers of databases are likely to require data dictionaries, which define the data fields and may even list the range of values for each field. Graphical representations of the system, possibly including predecessor systems may also be useful. For office automation files, folder structures, file classification plans and file naming conventions may also be important in terms of accompanying documentation. Finally transmittal information such as who prepared the records for transfer, how they were prepared, and who authorized the transfer is needed to maintain the authenticity of the records.

These five factors provide a sound basis for developing transfer policies. These factors emphasize the characteristics and needs of the records in terms of their reproducibility and the technological environment within which the preserver must operate. Perhaps most importantly the transfer of electronic records must be driven by preservation and

[3] Available from Sun Microsystems at http://www.openoffice.org/, checked 1 April 2004.

accessioning policies. The former, at least in the InterPARES context, explicitly sets out that the "process of digital preservation begins with the initial act of storage and extends through reproduction of the record."[4] The latter requirement, i.e., how the records are reproduced (presumably for access) is an important factor to consider at the time of transfer. At this time, how records will be accessed may be the greatest challenge to archivists as models for accessing electronic records are only just beginning to emerge. It is very possible that records can be reproduced in a preserver's lab environment, but not easily for providing access to institutional clients. That is to say, it may be possible to reproduce records so that their authenticity is evident to specialists, such as the archivists overseeing the transfer, but which is less evident to those not intimately familiar with the technological environment and transfer complications.

[4] InterPARES Preservation Task Force Report, p. 5, available at http://www.interpares.org/book/interpares_book_f_part3.pdf, checked 29 March 2004.