Preservation Policies and Strategies

Making Preservation Policies

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

In this session I want to pursue two main ideas. The first is a review of some of the fourteen points identified in the InterPARES Strategy Task Force Report. These points are designed to help frame policies for managing the preservation of electronic records. The second idea is an approach to the preservation of digital records proposed by the National Archives of Australia. As yet there have been no implementations of this proposed approach, as far as I know. I introduce it here because I feel it is a valuable addition in my own institution as a basis to developing policies and procedures particularly for electronic records involving some kind of functionality. Web documents are a good example: when is it sufficient to preserve a web page and when when does the record require the context of a functioning website to accurately convey its meaning? These principles can help re-orient the management of records in electronic format from a paradigm of media-centric records, where management is based on observable physical locations controlled by humans, to content-centric records management, where the management process is based on invisible logical locations controlled by computers."[1]

The InterPARES strategic framework is expressed in fourteen principles. In this way, experience and knowledge from a wide variety of institutions has been concentrated into general principles for broad consideration and application. I will highlight five of these principles.

Principle 3: recognize and provide for the fact that authenticity is most at risk when records are transmitted across space (i.e., when sent between persons, systems, or applications) or time (i.e., either when they are stored offline, or when the hardware or software used to process, communicate, or maintain them is upgraded or replaced). This principle applies to any existing model or strategy for the preservation of electronic records. No existing model assumes that technology will remain static and so whether custody is distributed, or whether emulation or migration are the preferred preservation strategies, records will move through space and time. In fact when records are reproduced, they are, in effect, moving through space and time every time the data is read from the storage medium (e.g., magnetic disk) and transmitted to the presentation medium (e.g., computer monitor).

If authenticity is heavily based on the identity and integrity of records, it is interesting to note that creators appear more concerned about protecting the integrity of the records

---

[1] Robert F. Williams, Cohasset Associates Inc. "Electronic Records Management Survey. A Call to Action" (White paper, 2004), p. 3. Available at http://www.merresource.com/whitepapers/survey.htm (checked 16 March 2004).

than identity.[2] This suggests to me that the identity component of authenticity is perceived by record preservers to be less at risk or less important than the integrity component.

Principle 4. recognize that preservation of authentic electronic records is a continuous process that begins with the process of records creation and whose purpose is to transmit authentic records across time and space.
This principle is consistent with findings from many other studies, including one by the ICA.[3] The ICA study considered the lifecycle to begin with the conception of the record, i.e., when the recordkeeping system is being designed. The InterPARES findings are based on the lifecycle beginning with record creation. This principle also requires that records appraisal be integrated into the continuous process of preservation.[4] Ongoing appraisal provides an ongoing basis for determining what is to be preserved, what is feasible to preserve, and perhaps most importantly identifies what documents actually have the qualities of a record.

It is also important to remember that in the InterPARES context the term "preservation" includes the ability to reproduce records. So the transmission of records across space and time is not simply for the preserver's maintenance of the records, but to reliably transmit them to the record user's environment as well.

Principle 5. be based on the concept of trust in records keeping and record preservation and specifically on the concepts of a trusted record-keeping system and the role of the preserver as a trusted custodian.
This principle captures a key ethical component of the preservation of electronic records. Because of the potential for undetectable manipulation of electronic records, a much less critical consideration in a paper technology environment, the preserver must be trusted and must have procedures in place that can be independently audited.

The implication of the chain of preservation model, which identifies that risks to the authenticity of records begin with the creation of the record, is that not only must the archival preservation processes be seen to be trustworthy, but so must the processes by which the records are created and maintained. A key motivation for the Sarbanes-Oxley law (which holds senior executives accountable for the accuracy of their records and reports) was the untrustworthy way in which financial accounting was managed in firms like Enron.

Principle 6. be predicated on the understanding that it is not possible to preserve an electronic record as a stored physical object: it is only possible to preserve the ability to reproduce the record.

---

[2] Heather MacNeil, "Providing Grounds for Trust II: The Findings of the Authenticity Task Force of InterPARES" in *Archivaria 54* (Fall 2002), p. 37.
[3] For example, see the ICA, Committee on Electronic Records, *Guide for Managing Electronic Records from an Archival Perspective* (ICA Study #8, 1997), p. 49-51.
[4] Principle 9: "integrate records appraisal in the continuous process of preservation".

This is the paradigm change identified by the InterPARES Preservation Task Force. It gets at the fundamental difference in terms of record keeping between paper and digital technologies. Preserving paper records is possible, because the paper medium is by comparison almost invulnerable to changes when it is transmitted across space and time. We can directly access it long after its creation.

It is worth considering this finding in relation to perhaps an even greater paradigm change: from oral or human memory record keeping to written record keeping. M.T. Clanchy, studied the transition in England from an oral to a written culture. Concerning the oral tradition he observes that "…no ancient custom could be proved to be older than the memory of the oldest living wise man. There was no conflict between past and present, between ancient precedents and present practice."[5] Contrast this with his statement about the new, paper-based recordkeeping environment: "Written records, on the other hand, do not die peacefully, as they retain a half life in archives and can be resurrected to inform, impress or mystify future generations."[6]

In the transition from paper to digital technology, there is ample evidence that the reliablity and durability of written records must be somehow maintained. That is, digital records must retain that "half life" Clanchy mentions and be retrievable for future needs. In the digital environment the transmissions through space and time that put records at risk happen much more frequently than is generally the case with paper records.

Principle 8. specify the requirements a copy of a record should satisfy to be considered equivalent to an original.
Because it is not possible to preserve electronic records, only the ability to reproduce them, it is essential that the requirements for an authentic reproduction be specified. With such a specification in hand, it becomes possible to measure the trustworthiness of a record. For example, a specific record can be authenticated long after its creation, assuming the relevant record requirements are known and somehow evident. Where the requirements are generally specified and evident, aggregations of records like series can be presumed authentic.

Making a Preservation Strategy

Currently there exist only two preservation strategies: migration and emulation.[7] Emulation allows preservation of the original format of the record. The record is accessed through emulating the behaviour of the original software and hardware on current technology. In this simulation of the records original environment the record can be reproduced. This strategy requires knowledgeable programmers to write the emulation programs[8] and might require record users to learn unfamiliar (i.e., obsolete)

---

[5] M.T. Clanchy. *From Memory to Written Record: England, 1066-1307* 2nd ed. (Oxford: Blackwell Publishers Ltd, 1993), 296.

[6] M.T. Clanchy, 296.

[7] It is my view that the Universal Virtual Computer initiative currently being undertaken is a variation on the emulation strategy. Similarly, encapsulation might be considered a sophisticated form of migration.

[8] Which in turn may require addressing complex intellectual property issues.

interfaces to access the records (e.g., DOS commands). An advantage to this strategy is that the record is reproduced exactly as it was in its original environment and thereby relieves the preserver from determining what characteristics of a record are essential to preserve. Migration requires changing the record format as often as needed over time to ensure that it continues to be accessible on the current technology. This approach requires resources for repetitive reformatting of holdings and the expectation that gradually accumulating losses to the record will occur over time. Two key advantages to this approach is that the preserver need only maintain current technologies, which is also the technology that record user's are likely to be most familiar with.

In 2002 Helen Heslop, Simon Davis, and Andrew Wilson, of the National Archives of Australia, wrote a paper entitled "An Approach to the Preservation of Digital Records". In it the authors outline an "approach to digital preservation that is being developed as part of the Agency to Researcher Digital Preservation Project."[9] It is worthwhile to consider the background to this paper before its substance and it is appropriate to do so in the context of this paper not only because it is one result of a significant change in approach at the NAA, but because of the significant change that all archival institutions will have to deal with at some point – which is how electronic records will remain accessible for use. As such it reflects a paradigm shift that is consistent, at least in my view, with the InterPARES findings, is interesting, but it perhaps takes one step beyond the steps of reconstituting and presenting the record as defined in InterPARES 1.

The approach outlined in this paper interests me because, on the basis of my own experience, it is possible to meet the preservation requirements outlined by InterPARES successfully, but without necessarily making the records easy to use. My feeling is that the approach outlined in the NAA's green paper outlines a means for planning effective ways of preserving the records so that how they are used is taken into consideration. For example, the acquisition of a large volume of correspondence (84 GB) exported for long-term preservation from a proprietary database format. Because the office has well-defined business and strong record keeping processes any given record within this acquisition are presumed to be authentic. In their operational environment, the database application maintained the records in a case file structure, that is, the scanned image of the incoming letter was connected to the word-processed response and any case notes completed each correspondence file. To preserve this critical file-level context, the records were exported in such a way that the linkages which established each file would also be preserved. In this way I believe we can preserve and reproduce these records with their authenticity intact. However, the process of reconstructing each file is a manual one. For an acquisition that encompasses tens of thousands of files, this effectively precludes the ability to work with large portions of these records. It is a question whether this is appropriate when one considers the way the records were used in their original environment. What is valuable in the NAA's "Approach" is that it provides a framework for consistently answering questions like that.

---

[9] Helen Heslop, Simon Davis, and Andrew Wilson. "An Approach to the Preservation of Digital Records" (2002), p. 6. Available at http://www.naa.gov.au/recordkeeping/er/digital_preservation/Green_Paper.pdf, checked 16 March 2004.

Like the InterPARES principles, the ones outlined in the green paper are also (or at least verge on being) ethical statements, that is guides to professional conduct. They also consider not just the ethical obligations to the records, but also the professional obligations to the record users and the preserving institutions.

The document begins by outlining a performance model for records. The model is very simple for paper records – using his or her eyes and hands a user interacts directly with a paper record. For electronic records two intervening steps between the user and the record are identified. These are the process and performance steps and relate to the record or source data file. The process step refers to the hardware and software required to correctly interpret the structure and format of the source record. The performance step refers to what is rendered as output (on a monitor's screen or through a printer, for example). The rendered record includes just those essential characteristics that qualify or affect the meaning of the record content. From this introduction, the paper's five principles are developed.

1. The digital preservation program must be able to preserve any digital record that is brought into custody regardless of the application or system it is from or data format it is stored in.[10]
This principle accepts an uneven record creation and maintenance capacity across the spectrum of possible donors of records to an archives. This is likely when the preserver's mandate includes the acquisition of records from private donors. It is very possible, however, that even a government environment will have a wide range of record creation and maintenance capacity. This principle also carries with it the requirements that the preserver have a consistent and effective resource base for preservation and that on the basis of this the preserver's capabilities be effectively integrated into the appraisal process.

2. The digital preservation program must determine and preserve the essence of the digital records in the National Archives' custody and recreate their essential performance over time.
The green paper's explanation of this principle indicates that
> determining the essence of a particular genre or type of record, such as a word-processed document or email, before the application of any preservation treatment is a very important way to ensure optimal, sustainable and accountable preservation.[11]

It was a task of the Authenticity Task Force to develop a typology of records for the purpose of identifying their authenticity specifications. The Task Force was not successful in this because
> the determination of documentary forms in general and the establishment of required elements of form and annotations in particular were deeply embedded

---

[10] The phrasing of the principles has been modified for the purposes of this paper by removing references specific to the National Archives of Australia.
[11] Heslop, Davis, and Wilson, "An Approach", p. 15.

within specific institutional and procedural contexts and were resistant to any easy generalizations.[12]

It would be interesting to see if studies of the NAA's record performance model would hold out greater hope for establishing a meaningful record typology.

The usefulness of a system typology, presented earlier, adds to my interest in this approach. It may also be that record performance will actually be system-based, that is it may be most appropriate to consider determining record performance on the basis of record aggregates. For example, an e-mail is generally considered to be a record. Is it more appropriate from an archival standpoint to consider the performance of corporate emails from an aggregate viewpoint, i.e., from the standpoint of all the emails of the corporation? If so, then the character or essence of the record moves from individualized messages similar to letters (to which they are often, if unfavourably, compared) to a transactional type of database – the essence or performance being defined by the email server. In any event, this principle is certainly consistent with the appraisal process requirements outlined by InterPARES.

The next two principles:
> 3. The digital preservation program will be based on non-proprietary technologies, and
> 4. To lessen the risk to the integrity of the records, the preservation program will minimise the number of preservation treatments applied to each digital record

address two fundamental ways to mitigate risks to the records. The final principle:
> 5. The digital preservation program will not limit the accessibility choices of the National Archives or of future researchers

is interesting because it addresses the obligations of the preserver to the researchers or clients, in particular in terms of the technological base that the user might require in order to access the record. This principle may be implied within the InterPARES principles. The explicit inclusion of the researcher's access in the five principles developed by the NAA is consistent with the integrated nature of the "approach" described in the paper and the ethical obligations of archivists to make records available.

These principles provide a strong basis for developing a preservation policy that addresses the commitments of the institution and the professionals employed at it, the commitments to the records to be preserved, and the commitments to the users of the preserved records. They build on and expand the four principles developed by the ICA's Committee on Electronic Records some years ago.[13] For institutions they define

---

[12] Heather MacNeil. "Providing Grounds for Trust II", p. 31.

[13] These are:
1. The archives should be involved in the entire life cycle of electronic systems…to ensure the creation and retention of electronic records that are authentic, reliable and preservable.
2. The archives should ensure that records creators create and retain records which are authentic, reliable, and preservable.
3. The archives must manage the appraisal process and exercise intellectual control over archival electronic records.

the purpose of preservation as a continuous process of transmitting records through space and time (IP4), that the preservation process must accommodate any digital record (A1); must be able to reproduce digital records to meet predefined specifications (IP6, A2); and must not rely on proprietary technologies (A3). Records must be reproduced as authentic copies and the risks to authentic records are recognized as resulting primarily through transmission through space and time (IP3); and thus as few preservation treatments as possible (A4). For users the preserver must be trusted to both acquire and maintain trustworthy records (IP5), but also to make them accessible as trustworthy records as transparently as possible (A5).

To conclude this presentation let me address briefly the actual definition and implementation of a preservation policy for electronic records within an institution. It is one thing to have a set of principles and understand how they might be used for developing an institutional preservation policy. It is another thing to actually to actually articulate that policy and, perhaps most importantly, have it implemented. What these principles require is a substantial change in what resources are required and how they are deployed. Consider, for example, the requirement to determine what constitutes the record, how its constituent parts can be preserved and appropriately assembled in correctly reproducing the record in a constantly changing digital environment. Achieving the first aspect requires developing appraisal processes that are applied throughout the life cycle of the record. At my institution, this requirement alone would mean that new procedures would need to be developed, time would have to be made available for staff to implement these new procedures, new types of records that would accurately and effectively convey appraisal information for subsequent appraisals within the record life cycle. The appraisal information would also have to be conveyed to those responsible for choosing technologies for record migration and assessing how effectively records are reproduced. Within my institution, not only are there no staff with the necessary technology skills, there is also no resource base currently to support the technology required to undertake it. For these reasons, the development of a preservation policy based on these principles will require strong and courageous leadership. Strong because of the competing demands on resources from established programs. Courageous because the resources required to maintain paper records will not decline, even though resources required to preserve electronic records will increase. Institutional heads will have to successfully request additional resources even while trust in the archival approach to preservation of electronic records is not yet established. Implementing the principles being articulated for preservation of authentic electronic records will require transformation within and outside our existing archival institutions. It may be useful to consider including a change management strategy in any preservation policy for electronic records.

---

4. The archives must articulate preservation and access requirements to ensure that archival electronic records remain available, accessible, and understandable.
ICA, Committee on Electronic Records. *Guide for Managing Electronic Records from an Archival Perspective* Study #8 (1997), p.8.