Framing the Preservation Paradigm: A Report on the Findings of the InterPARES Project

Jim Suderman

Abstract

The presentation will summarize the InterPARES research by considering how the digital environment has affected two central concepts: the records lifecycle and the idea of record. Specifically, the presentation will consider the importance of the records lifecycle model from the perspective of creator and preserver by relating it to the InterPARES Manage Chain of Preservation model. The presentation will also consider the nature of electronic records as conceptual entities with particular emphasis on authenticity. The implications of these concepts for archival appraisal and preservation will be examined in relation to existing theories, best practices, and legislation.

Archivists are in the time business. A glance at the ICA's *Code of Ethics* confirms this. The very first item in the C*ode* requires archivists to "protect the integrity of archival material and thus guarantee that it continues to be reliable evidence of the past."[1] Protecting reliable evidence of the past may be the most obvious aspect of the time business that we are involved in. In my presentation today I will explore this aspect of time as well as two other related considerations. These other two include the records lifecycle and some ways that the digital medium has affected the archival activities of appraisal and preservation. This will be an effective structure for sharing the concepts and emerging findings of the InterPARES Research Project.

The latin words "Inter PARES" mean "among peers". As an acronym it stands for International Research on Permanent Authentic Records in Electronic Systems. The InterPARES research project has been underway since 1998, developing theoretical and methodological knowledge to enable the preservation of trustworthy electronic records over the long-term. For this reason, the first phase of the research necessarily took the perspective of the preserver. The kind of records focussed on during the first phase were those at the inactive stage of their lifecycle. Generally these records had some kind of analogy to the kinds of records that exist in paper form. The second phase of the research, currently underway and a little over a year away from completion, takes the perspective of the record creator. This phase is focussing on records that are still active within the record creator's context. The current research also extends beyond administrative records such as those found in government, to include records from the artistic and scientific communities. This perspective enlarges the findings of the first

[1] International Council of Archives. *Code of Ethics* (1996), accessible at http://www.ica.org/biblio/code_ethics_eng.html.

phase of the research by considering how records become authentic and trustworthy in the first place.

Consistent with the enlarged scope of the second phase research is the increased number of researchers and range of disciplines involved. There are approximately a hundred researchers, from twenty countries, connected with the current, second phase of InterPARES. They are divided between record creators (artists, scientists, scholars of the arts and sciences), technology experts, and record experts (archivists, records managers, information managers). The earlier phase had sixty researchers, predominantly archival scholars and archivists with a few computer scientists and lawyers, from a total of thirteen countries. Through its participants, InterPARES research brings together with diplomatics and archival science many fields of study, such as music and geomatics,. The involvement of academics from a range of artistic and scientific areas provide an interdisciplinary approach that extends beyond what was possible in the first phase of the research. Case studies and surveys form the raw material for analysis. Completed case studies have examined record creation in a range of environments, such as the creation of a work of music or a digital atlas or land registry records. The scope of surveys is similarly broad and examples include the recordkeeping practices of digital photographers or studies of government website interactivity. Process models are created and diplomatic analyses undertaken on the basis of these studies and surveys.[2]

Through my presentation I will relate some of the concepts and findings emerging from the InterPARES research to archival theory and practice, particularly as it relates to appraisal and preservation. Some of the concepts being reviewed may seem elementary and I ask for your patience. These concepts are reviewed here because the digital environment has proven a

[2] The methodological principles of the InterPARES 2 Project are available at http://www.interpares.org/ip2/ip2_methodological_principles.cfm.

challenge to the most basic assumptions within archives, notably the definition of what a record really is – the centre of the archival endeavour! Before talking about the idea or concept of record I will explore the records lifecycle. The theory of the records lifecycle is central to InterPARES research because it identifies two complementary sets of responsibilities relating to the record, those of the creator and the preserver.

The lifecycle concept is a fairly recent one, dating roughly from the middle of the last century. Prior to this records were either "live," that is participating in the activity to which their creation was related, or "dead," meaning that the activity to which they were related was wholly concluded. Beginning in the 1950's, Canadian and American records managers – records managers being agents of the creating organization – were identifying three lifecycle phases, based on frequency of use. The first phase included current records because they were still actively and frequently used. As the record creator's requirements for the records diminished they continued to be active, but this period of reduced activity was called semi-current or semi-active. This second phase frequently coincided with storage of the records in an off-site facility, meaning that they were not immediately at hand but had to be retrieved. This inefficiency was tolerated because of the infrequency of use. The third phase occurred when the creating organization perceived no further requirement for the record. Records in this phase were referred to as non-current or inactive. According to authorized dispositions, inactive records were either destroyed or transferred to a preserver, such as an archives, for permanent preservation. The relationship of these phases to physical locations continues to be reflected in existing legislation in both France and Italy (and possibly other jurisdictions as well). Meanwhile the theory of "movable responsibility" was being developed in the United Kingdom. This theory recognized the joint responsibility of records managers and archivists to administer

records. The theory expressed the gradually diminishing responsibility of the records manager over records as they move through the lifecycle while the responsibility of the archivist for preservation and access increases. The significance of this shifting responsibility was the recognition that records, used and relied upon by the creating organization throughout their active period, must not be diminished in any critical way as they make the transition to the archival preserver.

The InterPARES 2 definition of the records lifecycle is:

> The theory that records go through distinct phases, including creation or receipt, use and maintenance, and disposition (destruction or permanent preservation).[3]

It is possible to discern two kinds of time in the records lifecycle. The first sort is connected with the creation and use phases. At this stage of the lifecycle, time is defined in relation to the creator's administrative requirements. These are derived from the creator's activities and the juridical environment in which those activities take place. Most obviously time is reflected through specifying retention periods for the records. The creator assesses the amount of time needed to retain the records and that period is the use and maintenance phase for those records. The second kind of time begins when records reach the end of their retention period. At the disposition phase they are either destroyed or transferred to a preserver. In the preserver's realm the records enter a kind of atemporal existence, that is, time passes but the preserved record is unaffected by its passage to the extent consistent with the preserver's capacity.

The theory of a shared and shifting responsibility between the record creator and preserver was confirmed within the InterPARES research on electronic records. To meet the creator's requirements, electronic records are produced and may be modified or converted due to

[3] InterPARES 2 Terminology database.

changing technology. If records are converted or modified in some way to meet changing technology but they continue to be maintained by the creating organization for further action or reference, then the records remain authentic. Transferring records to the preserver constitutes a change in status of the records. They are no longer maintained for action or reference in the daily affairs of the organization, but are preserved. In other words, the purpose for their existence has changed. So while the creator may have modified the records by changing their form or the system that supports them, as long as the records continue to be related to the activities of the creator, and relied upon by the creator, then the authenticity of the records remains intact. On the other hand, it is possible, even likely, that electronic records will be modified by the preserver. But the preserver's purpose in making modifications is not to support the on-going business of the creating institution, but rather to maintain the records as authentic copies of the records of the creator. This change in purpose, from the creator's creation and manipulation of them for their support in the daily operations of the organization to the preserver's responsibility to maintain them as authentic copies of the organization's activities, constitutes the substance of the two roles within the records lifecycle.

A factor that complicates this conclusion is found in the laws that govern records. Although laws commonly reflect the records lifecycle in their provisions, most fail to do so consistently or comprehensively. For example, an e-government law may outline the requirement to authenticate and complete a record with a digital signature – which is a record creation requirement. However, since digital signatures cannot be preserved, such a legal requirement fails to consider the feasibility of maintaining or preserving the complete record over the long-term. This sort of inconsistency may be due to assumptions from the pre-digital

environment influencing legislation specifically targetted towards the current, digital environment.

Paper and ink, as a record-making technology, have a durability that normally extends long beyond the creation and maintenance phases of records. And so, in the pre-digital environment it was, generally speaking, only in the archival field where the technological lifecycle of paper became a consideration. In response to the inevitable deterioration of paper over decades or centuries, conservation labs were established and approaches and procedures for carrying the records forward, beyond the usual lifecycle of the paper technology, were implemented.

One of the significant differences with the digital environment of course is that the durability of the computer technology used to create electronic records has a much shorter lifecycle than does paper and ink. So much shorter that intervention is frequently required to sustain records while they are still in active use. The effect of this has been to make the shorter technological lifecycle visible and a concern to people other than just archivists and conservators. The digital lifecycle is driven by technological innovation as well as market forces, I think it is fair to say. By contrast, the records lifecycle is driven by legal and administrative considerations. The question arises that if the technological lifecycle is shorter than that of the records, and if it is also not primarily governed by the creator's own needs and requirements, then what will be the effect of changing technology on the records? Archivists are concerned because if the technology supporting the records is changing while the records are still active, then there is little point in appraising them at the end of that period.

One final observation concerning the records lifecycle is that within the digital environment there appear to be fewer situations where records *naturally* move through the

different phases. That is, records can be created but are never set aside within a recordkeeping system. In the paper environment the physical bulk of the records ensured their eventual removal to semi-active storage and from there to destruction or a preservation repository. While the growth of information in digital form is frequently noted, this growth is matched by growth in digital storage capacity combined with an ever-increasing miniaturization of storage devices. The effect of this is a diminishing of the naturally occurring physical imperative for records to move through their lifecycle that occurs with paper records. There is also no legal requirement for records to move through a lifecycle. If records do not naturally move through their lifecycle that will affect the shift or transfer of responsibilities from the record creator to the preserver.

A key InterPARES response to this situation is the development of a "whole of lifecycle" model that reflects the perspectives of both the creator and the preserver. This is called the Manage Chain of Preservation (MCP) model.[4] Although it is not a process model, the MCP model provides a comprehensive yet, I believe, flexible framework that accommodates varying legal requirements of different jurisdictions and that helps ensure that records are created, appraised, and preserved as required.

Included within the MCP model are descriptions of three systems or sets of rules governing records. The *record-making* system addresses the development of record forms and record-making metadata schemas. It also establishes access privileges, integrates business procedures with documentary ones, and helps determine technological requirements for record-making. The second system is the *recordkeeping* system. It encompasses registration and classification schemes, retention schedules, procedures and technology for maintaining authentic records and a retrieval system to enable the creator to use the records for action or reference. The third system is for *permanent preservation.* This system incorporates systems for appraisal

and description and for reproducing the records for access.[5] The preservation system corresponds to the preserver's responsibilities as envisioned by InterPARES, which is where records are separated into their various digital components, reformatted for the preserver's technological environment, etc. Recall that what distinguishes the preservation system from the other two, i.e., record-making and -keeping systems, is that actions taken by the preserver within this system are for the purposes of preservation, not for the records creator's purposes of supporting business activities.

The model's name, Manage Chain of Preservation, embodies a significant difference from traditional archival appraisal practice. Pre-electronic record creation generally resulted in physical records that were difficult to change, and if changes were made, it was difficult to hide those changes. For this reason archivists did not normally need to consider the authenticity of the records in the pre-electronic environment during appraisal, an activity normally undertaken late in the lifecycle. As long as they were received from the record creator, their authenticity as the records of that organization could be presumed. All that had to be managed in the paper environment was the chain of custody to ensure that records moved directly from the record creator to the preserver. Because of the capabilities of computer technology, the preserver can no longer simply presume that records received from the creator are authentic. Therefore, the preserver must know that record authenticity, in terms of their identity and integrity, is established when they are created and maintained throughout the lifecycle so that they can be transferred to the preserver with authenticity intact. Assessing record authenticity therefore becomes an appraisal activity and a central reason for integrating record creation with preservation within the Manage Chain of Preservation model.

---

[4] See http://www.interpares.org/ip2/ip2_models.cfm.
[5] System descriptions are taken from the MCP Model Activity descriptions, available at link in note 4.

Naturally, before the authenticity of a record can be assessed, one has to know what the record is in fact. The InterPARES definition of an electronic record is consistent with the traditional archival definition of record: "a document made or received in the course of a practical activity as an instrument or a by-product of such activity, and set aside [in electronic format] for action or reference."[6] This definition sets records apart from the more inclusive category of documents, which are not instruments or by-products of practical activities nor are they set aside. Information is simply a message intended for communication across space or time. Data are the smallest meaningful pieces of information.

As with the records lifecycle, however, the definitions of records differ, sometimes considerably, among the laws and statutes within and across jurisdictions. Many laws that relate to records do not relate them *to activities* or require that they be *set aside* for future action or reference. For example, laws that enable archival institutions frequently refer to records simply as commodities or artefacts that can be or must be transferred. The one consistent exception within the haphazard definitions of records within statutes was found in evidence law, which emphasizes the importance of the relationship of the record to the event both in terms of content and time of creation.

Following on what was observed about legislation earlier in connection with the records lifecycle, some laws, particularly those supporting electronic commerce or electronic government, authorize or even require the addition of digital signatures to authenticate records in transmission. In some cases however the digital signatures have an on-going task to authenticate records that have been stored and set aside. Such laws require records to be created with a component that is governed by a technological lifecycle rather than one dictated by the business process to which the records are connected. Good practice requires changing digital signatures,

[6] Glossary definition, InterPARES 2 terminology database.

including the encryption algorithms they use, usually within two years or so of their being issued to an individual. If the record must be maintained for ten years this imposes an impossible burden on the record-keeping system. It is impractical to renew the digital signatures and their security function will be compromised by emerging technologies. As a result, laws calling for the addition of digital signatures as an integral component of a record, for example in replacement of hand-written signatures, require the creation of records that cannot be preserved.

InterPARES has used the science of diplomatics to identify the characteristics of electronic records. This science of course long predates the digital era but the IntePARES research illustrates that the principles it espouses can be effectively applied in the electronic environment. The digital environment, in which there is no tangible artefact to help comprehend and give shape to the record, requires diplomatic science (or some equivalent) to establish what records there are and where they are found within systems. Using diplomatics the following characteristics of an e-record were established:

1. A fixed form – this does not mean that the binary encoding of the electronic record cannot change, just that when the record is rendered or reproduced it does so with the same documentary form as when it was first completed.
2. Stable content – the message must not change which in the electronic environment means that appropriate security measures need to be in place to ensure that only authorized individuals create content and that once created it cannot be modified or at least not without there being a record of the modification.
3. An action – to which the record is an instrument or a by-product of the organizational procedures followed;

4. Explicit linkages to other records pertaining to the same activity, for example as provided by a file classification code;
5. An identifiable administrative context;
6. An author, addressee and a writer.

The case studies developed within InterPARES provided the primary basis for determining how these characteristics manifested themselves within electronic records and the systems used to create and maintain them.

In a paper technology environment the record attributes that might express the characteristics outlined above are easy to imagine. For example, the signature or the letterhead on the paper may indicate the author of a letter. Other characteristics may be external to the record, such as the identifiable administrative context which may take the form of a procedure manual. In the electronic environment these characteristics may not be expressed within the documentary form of the record. For example, the action to which a record pertains and linkages to other records pertaining to the same activity may exist in a tracking database external to the record. Thus the value of a record is closely tied to the availability of this kind of metadata, which may be defined here as data expressing the attributes of a given record.

Some attributes may also manifest themselves as separate digital components. Digital components are simply technological containers for storing data. Other attributes may exist within the operation of the software, and so on. To illustrate this idea, consider a report in electronic format that contains a picture. The text of the report is one digital component while the picture is a second one. Both components are part of the content of the report and must be combined within the system to produce the record. It may be through the software, a web browser application for example, that the two components are correctly displayed.

Because the attributes of electronic records may be located in various digital components, InterPARES concluded that appraisal needs to occur at the point of record creation. Without determining at the outset what the records are and what components are required to produce them there is no way that the preserver, late in the records lifecycle, can assert that the records being preserved are those created and used at the beginning of the lifecycle. That is, the preserver cannot assert that the records received from the creator are authentic. Appraisal early in the lifecycle is not especially new nor unique to InterPARES and may be said to be generally accepted within the archival community. What is frequently not emphasized is the activity of monitoring the records following early appraisal. The purpose of monitoring is to ensure that, when the inevitable technological changes occur during the maintenance of the records, that the substantive attributes of those records remain unaffected. That is, the relationship of the record components that comprise the record, the relationship of the records with other records, and their relationship to the business activity or activities which brought them into existence. If monitoring determines that substantive changes are being or have been made, then the original appraisal conclusions must be revisited. Both the initial appraisal and the monitoring of the records provide a means to establish the trustworthiness of the preserver of the records. If the preserver, through the appraisal and preservation processes, is to provide the warranty of the authenticity and reliability of the records through time, then it is essential that the trustworthiness of the preserver can be demonstrated.

To support the appraisal and monitoring of records, seven benchmark requirements were identified within InterPARES.[7] For the purpose of this presentation I will only outline the first of these, record identity and integrity. Identity is established through the names of the

[7] See "Requirements for Assessing and Maintaining the Authenticity of Electronic Records", at http://www.interpares.org/book/interpares_book_k_app02.pdf.

individuals connected to the record's formation (author, addressee, etc.). Also required are the date of creation or transmission, the activity to which the record is connected, and the record's relationship to other records. Integrity is established if the record is both complete and its message unmodified. For example, a memo where the text was complete and but another attribute, e.g., the author, was absent would indicate that the integrity of the record was compromised.

InterPARES further concluded that appraisal at the creation stage of the records lifecycle is not simply to assess the authenticity of the records but also to determine or estimate whether it is within the means of the preserver to preserve the record. Feasibility of preservation has not generally been a consideration in the pre-digital environment but ensuring access to all the digital components of the records and ensuring that the relationship of each component to the others can be maintained requires significant technological resources and skills. These may not be available within every archival institution, necessitating alternatives such as

- Postponing the acquisition of the records until increased resources are available;
- influencing how the records are created, i.e., create records so that it is feasible for the designated preserver to preserve them with its existing resources;
- sharing resources with other preservation institutions;
- preserving modified copies of the records – in the example above where it may not be feasible even for the best resourced institution to preserve records with digital signatures, the best option may be to preserve the records without that component, noting that fact for future users of the records, of course.

InterPARES research has looked beyond authenticity to consider the qualities of reliability and accuracy as well. These have a narrower scope within the InterPARES research

because, unlike authenticity, reliability and accuracy are qualities that fall exclusively within the responsibility of the record creator. A reliable record is defined as "one whose contents can be trusted as a full and accurate representation of the transactions, activities, or facts to which they attest and can be depended upon in the course of subsequent transactions or activites."[8] Accuracy is a very similar quality and is familiar to InterPARES primarily in connection with scientific records and databases. It is therefore assessed in terms of correct and exact content. The fact that these two qualities are wholly within the scope of the record creator means that the preserver need focus solely on record authenticity.

By now the role of metadata, i.e., data that expresses attributes and other details about the records, is clear although it has not been central to my discussion. There are a number of recordkeeping metadata schemas in existence and it is not the intention of InterPARES to produce another one. However these schemas are being inventoried and analyzed into MADRAS (Metadata and Archival Description Registry and Analysis System).[9] Thirty-three schemas have been identified and the analysis involves assessing how well they meet recordkeeping standards including:

- the authenticity requirements identified in the first phase of InterPARES,
- the ISO Recordkeeping metadata principles (ISO 23081), and
- the Australian Record-Keeping Metadata Standard for Commonwealth Agencies.

Findings from this research may help organizations determine which schema best meet its record keeping needs particularly with regard to establishing and maintaining the identity and integrity of records.

[8] Definition taken from the ISO 15489 Records management standard, section 7.2.3 "Reliability".
[9] See http://www.gseis.ucla.edu/us-interpares/madras/index.php.

Regarding preservation, the conclusion reached during the first phase of the Project was that it was not possible to *preserve* electronic records. The nature of the technology is such that the bitstream, literally the 1's and 0's that make up the digital components of which an electronic record is comprised, will change. But that does not necessarily mean that the record cannot continue to perform the function it was created to perform. For example, a digital image in one format, such as .tiff, will have a different bitstream from the same image in another format, e.g., jpeg. To the computer the two entities are completely different. But to a human, both may serve the purpose for which the image was created. So although it was concluded that it was impossible to *preserve* electronic records, the technology does allows records *to be reproduced*. This is not so different, in fact, from the paper environment. Archivists routinely take paper records out of their physical context, cull extraneous materials, put them in boxes in vaults, and require researchers to come to archival reading rooms to use them. But these activities tend to affect the context of bodies or aggregations of records. With electronic records, archival processes, including the transfer to the preserver's system, may well have an impact at the sub-record level or within the record. That is, elements of record content may reside in different computers, records may have to call supporting computer files to provide the structure for the data comprising the record content. For this reason, determining the characteristics of the records under review is essential, so that they, along with their attributes can be reproduced.

We have already seen the implication of this finding. Record characteristics and attributes have to be determined within a fairly short period of time. They must be identified while the records were active and before their underlying platform had changed in ways that actually changed the records or their attributes. This means applying the process of appraisal to active records specifically to assess and support their authenticity. Not doing so would risk

preserving records in ways that might result in their loss of authenticity. Furthermore the *creator* might maintain records that are *still active* in such a way as to put their authenticity at risk. Using appraisal to define records in terms of their content, characteristics, and attributes, has a second benefit. Through appraisal the preserver can also determine if it is feasible to preserve the records. This is a key conclusion as it speaks to the ethical obligation of archivists, outlined at the outset of my presentation, to preserve the authenticity of the records entrusted to their care.

Archival description is perceived within the InterPARES framework as a key preservation tool. It has a different role from that of the record metadata. In its traditional role, archival description provides the warranty of authenticity for the records being described. It does this through communicating the relationship of the records with the activities to which they adhere and explaining the nature of the relationships with other records. In the electronic environment description becomes the common thread holding records together as they undergo continual technological change. Archival description goes beyond establishing the identity and integrity of individual records to provide a comprehensive authentication of the whole body of the records being described. This is achieved by representing that body of records as one that must emerge from the creating organization in its usual and ordinary course of business. That is, it authenticates the relationship of the creating organization with its on-going business activities.

InterPARES research has been underway for almost seven years by now and its findings are beginning to take shape. The research is expected to produce a framework of record keeping principles and criteria to help organizations develop whole of lifecycle record keeping policies for the electronic environment. Anticipating its findings somewhat, principles for records creators may advise creation of records that are feasible for the preserving agency to effectively

preserve. That is, defining records in such a way that supports only the records lifecycle but also the technological lifecycle of the digital components of which the records are comprised. Another principle may be establishing systems with explicit events or triggers that move records through their lifecycle. For the preserver, principles may include not only advising appraisal at the point of record creation specifically to assess the authenticity of the electronic records and monitor them throughout their lifecycle. Feasibility considerations in connection with accountability requirements will assist legislators to craft laws that enable archival programs to fulfil their accountability role within democratic societies. The MCP model, already available, provides a comprehensive framework for integrating lifecycle considerations that may be inconsistent or expressed in a piecemeal fashion within national statutes or organizational by-laws.

The challenge for archivists will be to move their work as preservers forward in the records lifecycle. This needs to be accomplished without compromising the impartiality of the records being created as instruments and by-products of the activities of the time and place to which they are connected. Let me close with a quotation from Dr. Duranti, the Project director:

> InterPARES is a long way from its conclusion and much knowledge is still in the course of being developed…What is not going to change are a few points of principle that InterPARES has been able to prove time after time, that is: first, technology cannot determine the solution to the permanent preservation of electronic records; second, archival needs must define the problems and archival theory must establish the correctness and adequacy of each technical solution, and, third, solutions to the preservation problem are inherently dynamic, thus ongoing research is vital to enable archivists to deal with the challenges presented by the new information technologies.[10]

[10] Unpublished paper presented at Havana Colloquium, 16 March 2005.