

Continuity and Transformation in the Role of the Archivist: The Findings of the InterPARES Project

Luciana Duranti

Since its inception in 1998, the InterPARES project has had the goal of developing the theoretical and methodological knowledge that will enable archivists to preserve electronic records over the long-term in such a way that their trustworthiness can be maintained intact. The first phase of InterPARES took the perspective of the preserver, focused on inactive records, and addressed digital records analogous to those on paper and maintained in document management systems and databases. These records were mostly textual and required by the creator for accountability purposes. The second phase of InterPARES, which began in 2002, took the perspective of the creator, focused on active records, and addressed digital records produced in dynamic, interactive and experiential environments in the context of artistic, scientific, and e-government activities.

Both the composition of the research team and the methodologies involved in the two phases were dramatically different. Most researchers in InterPARES 1 were either archival scholars or archivists, joined by a few computer scientists and lawyers, while InterPARES 2 researchers were evenly divided in three groups: records creators (i.e., artists, scientists, administrators, and scholars of arts and sciences), records preservers (i.e., information and records managers, archivists, conservators, and archival and conservation scholars), and technology experts (i.e., computer scientists and engineers, and scholars of knowledge organization, information systems, multimedia, computer science and engineering).

The methodology of InterPARES 1 had at its core archival and diplomatic analysis of case studies selected on the basis of grounded theory, that is, by changing the selection criteria according to the findings of each round of cases. Diplomatic analysis was developed to guide the management of contemporary records from the moment of their creation. In contrast, InterPARES 2 relied on the principles of interdisciplinarity, transferability of results from one disciplinary field to another, open inquiry, and multi-method design, thereby adopting a variety of methodologies derived from all the disciplines involved, as needed by each research activity.¹

Finally, InterPARES 1 comprised sixty researchers from thirteen countries, while InterPARES 2 comprised one hundred researchers from twenty-one countries. Ironically, the greater variety of cultures, disciplines, methodologies and perspectives within InterPARES 2 has facilitated the development of hypotheses, their analysis and the achievement of substantial results. Although, knowing how difficult it is to find any agreement on theory and methods within the archival community, the fact that it is easier to do so across professional and disciplinary communities should not be so surprising at all.

The InterPARES 1 and 2 research has produced a large amount of findings and products. I will focus on two areas: the concept of record and the method of selection.

¹ See "Overview of Intellectual Framework" posted on the public area of the InterPARES 2 website, at www.interpares.org/InterPARES2/methodology/

The InterPARES research team determined at the outset that it was necessary to establish the nature of the object of this study, at least as a hypothesis that the actual investigation could support or overturn. In other words, the team needed to agree on the definition of record and on how it differed from document, information and data. It decided to adopt the traditional archival definition, which states that a **record** is any document created (i.e., made or received and set aside for action or reference) by a physical or organizational person in the course of a practical activity as an instrument and by-product of it. Then it proceeded to define **document** as information affixed to a medium in a determined form, **information** as a message intended for communication across space or time, and **data** as the smallest meaningful piece of information. Finally, it defined an **electronic record** as a record created (i.e., made or received and set aside for action or reference) in electronic form, meaning that a message received in electronic form but set aside (filed) in paper form is a paper record, while a letter received on paper but scanned in the computer and only used as a digital file is an electronic record.

However, to understand the concept of electronic record implied by the archival definition, that is, to be able to identify the necessary characteristics of such a record, the team decided to use diplomatic analysis. As a result, the following characteristics of a record were identified: 1) a fixed form, meaning that the binary content of the record must be stored so that its message can be rendered with the same documentary form it had when first set aside; 2) a stable content; 3) explicit linkages to other records within or outside the digital system through a classification code or a naming convention; 4) an identifiable administrative context; 5) an author, an addressee, and a writer; and 6) an action, in which the record participates or which the record supports either procedurally or as part of the decision making process.

Like traditional records, electronic records have formal elements, which might be extrinsic or intrinsic, but in addition to them, they also have **attributes**, which are the defining characteristics of each given record or of a record element in it. An attribute may manifest itself as one or more elements of form. For example, the name of the author of a record is an attribute, which may be expressed as a letterhead or a signature, both of which are intrinsic elements of documentary form, that is, record elements. In addition to attributes that manifest themselves in the form of the record, that is, on the face of the record, as record elements, every record has attributes that are implicit in other parts of the record, such as the name of the creator or the name of the medium, but in digital records they are also expressed, albeit outside the documentary form. Because of this, they are mostly transparent to the user, and manifest themselves as metadata included in either a record profile,² a topic map, or other digital entity linked to the record.

Finally, with electronic records, we also have to differentiate elements and attributes from the record's digital components. A **digital component** is a digital object that may contain all or part of a record, and/or the related metadata, or more than one

² A record profile is an annotation inextricably linked to the record that includes several fields, which are either automatically or manually filled in with the record's metadata.

record, and that requires specific methods for preservation.³ In other words, a digital component is a unit of storage.

If these were fundamental findings of InterPARES 1, InterPARES 2 found that in interactive and dynamic environments records have additional characteristics and the concept itself is much more nuanced. In such environments, a key distinction is that between stored records and manifested records. A **stored record** comprises the digital component(s) used in re-producing one or more than one record, which include the data to be processed in order to manifest the record (content data and form data) and the rules for processing the data, including those enabling variations (composition data). A **manifested record** is the visualization or materialization of the record in a form suitable for presentation to a person or system. Sometimes, it does not have a corresponding stored record, but is re-created from fixed content data when a user's action associates them with specific form data and composition data (e.g. a record produced from a relational database).

InterPARES 2 also found that there are several types of electronic records, which can be divided in two fundamental categories: static and interactive. **Static records** are those that do not provide possibilities for changing their manifest content or form beyond opening, closing and navigating. Examples are e-mail, reports, sound recordings, motion video, snapshots of web pages. **Interactive** records present variable content, form, or both, and the rules governing the content and form of presentation may be either fixed or variable. They comprise non-dynamic and dynamic entities. **Non-dynamic entities** exist when the rules governing the presentation of content and form do not vary, and the content presented each time is selected from a fixed store of data. Ex. Interactive web pages, online catalogues, records enabling performances—they are indeed records when all the other conditions for the existence of a record are satisfied. **Dynamic entities** exist when the rules governing the presentation of content and form may vary—they are either records or potential records.

They are **interactive records** when they have fixed form, stable content, or a bounded variability. In an interactive environment an entity has *fixed form* if its binary content is stored so that the message it conveys can be rendered with the same documentary presentation it had on the screen when first saved (different digital presentation: Word to .pdf); and if the same content can be presented on the screen in several different ways in a limited series of possibilities: in such a case we have a different documentary presentation of the same stored record having stable content and fixed form (e.g. statistical data viewed as a pie chart, a bar chart, or a table). An entity has *stable content* if the data and the message are unchanged and unchangeable, meaning that data cannot be overwritten, altered, deleted or added to. An entity has *bounded variability* when changes to the form are limited and controlled by fixed rules, so that the same query or interaction always generates the same result, and we have different views of different subsets of content, due to the intention of the author or to different operating systems or applications.

They are **potential interactive records** when the variation is due to data that change frequently (the design permits updating, replacement or alterations; allows data collection from users or about user interactions or actions; or uses these data to

³ See Preservation Task Force Report in *The Long-term Preservation of Electronic Records: the InterPARES Project*, cit. <<http://www.interpares.org/book/index.cfm>>.

determine subsequent presentations), and when the variation is due to data received from external sources and not stored within the system (see InterPARES case study VanMap).⁴ We also have potential interactive records when we have entities produced in dynamic computing applications that select different sets of rules to produce documents, depending on user input, sources of content data, and characteristic of content (e.g. weather sites), and when we have entities produced by evolutionary computing where the software generating them can change autonomously (e.g. scheduling and modeling of financial markets; edutainment sites).

InterPARES 2 also found that records in interactive and dynamic environment have additional functions with respect to traditional records. We know from diplomatics that traditional records can be dispositive, probative, supporting or narrative, but these new records types can have also instructive or enabling functions. Records have an **instructive function** if they delineate the form in which external data are to be presented (e.g., scores, scripts, regulations, manuals of procedure, instructions for filling out forms); while they have an **enabling function** if they enable performance of artworks (software patches), execution of business transactions (interacting business applications), conduct of experiments (a workflow generated and used to carry out an experiment of which it is instrument, byproduct and residue), analysis of observational data (interpreting software), etc.

A full discussion of these constructs can be found in an article written by Luciana Duranti and Ken Thibodeau.⁵ However, I introduced here these few concepts in order to link them to the methodology for carrying out the function of appraisal as it has been elaborated by InterPARES 1 and 2.

When thinking about appraisal, the central question with which we are confronted is whether appraisal as we understood it in the paper world has changed when confronted with the digital environment. It is easier to answer this question if we make a clear distinction between selection and appraisal.

Selection of records is the process that starts at creation and continues till the records are acquired by the archives for permanent retention, and is therefore the prelude to preservation, increasingly part of preservation itself. **Appraisal**—which is an attribution of value to records, a value external to the records and defined by the acquiring institution or program—is supposed to justify the choice of continuing retention for a possible permanent preservation. But, let's talk first about **selection**.

With digital records, the function of selection has changed in four fundamental ways: first, the preserver must assess the **authenticity** of the records considered of continuing value; second, the preserver must determine the **feasibility** of the preservation of the authenticity of the records; third, the disposal decision must be made very **early** in the life of the records; and, fourth, the preserver must constantly **monitor** the records of the creator and, if warranted by the changes that they have undergone through time, revise the disposal decision.

Are these changes of a methodological or of a theoretical nature? In other words, is the selection process required for digital records a different way of doing the same

⁴ See final report on the case study at http://www.interpares.org/display_file.cfm?doc=ip2_cs24_final_report.pdf.

⁵ Luciana Duranti and Kenneth Thibodeau, "The Concept of Record in Interactive, Experiential and Dynamic Environments: the View of InterPARES," *Archival Science* 6 (2006):13-68.

thing or does it reflect a radical alteration of the nature of the selection function? I would say that the former statement is true and that the theory of selection and, within it, the theory of appraisal, have not changed: their purpose is still that of preserving a concentrated, essential record of the past that maintains intact all its archival characteristics of naturalness, impartiality, interrelatedness, uniqueness, and, most importantly, authenticity, without which records could not be trusted either as evidence or as sources of information.

But, if not the selection theory, something quite essential to fulfil the requirements of such theory has changed. To start with, the **methodological requirements** posed by the digital environment are quite radical and require that archivists be involved with the creators on an ongoing basis and walk the very thin line that separates a competent guidance to proper creation and maintenance procedures from outright interference in the making and keeping of the records. It is a difficult balancing act, vital to the protection of the impartiality of the records, but it is possible to accomplish it if organizational policies, strategies and plans of action exist to support the work of the archivist. Indeed the "external" values that guide the appraisal of records in the course of the selection process must come from those written policies, strategies and plans of action. Much more rigour is required with digital records than ever was with paper records, because of the trust that is put in the designated preserver, the only possible neutral third party handling materials that are easy to manipulate, tamper with, accidentally corrupt, or destroy.

In addition, the **scalability** of the process is an issue. While all archives will eventually acquire digital records, and none is completely ready for this, most archives do not have the resources to even begin to get ready. Thus, it is important to scale down the selection process in terms of basic requirements and to identify in each specific context what is essential and what is not for taking control of the digital records under the jurisdiction of each given archives. (This is indeed the goal of the third phase of the InterPARES project, implementing solutions in small and medium or low resources archival organizations or programs, which I will discuss later).

Most importantly, the selection process I just described is only relevant and possible with digital records that are created by organizations, are kept in record keeping systems or in document management systems, are subject to laws and regulations, formally participate in procedures, etc. Even so, there are big obstacles to implementation also in these situations, and they derive from the quality of the records, from their poor metadata and documentation, from the archives ability to maintain their characteristics, including functionality and interoperability, and from the limits of the human and financial resources available. This might mean that, before examining and assessing the value of records, we have to focus on records creators and their functions, and learn how to appraise records by understanding functions, conducting simultaneously a top down and a bottom up analysis. It might also mean that appraisal needs to become an iterative process that only stops when the records acquired by the archives are designated as destined to permanent preservation.

But the most significant change I see has to do with the object of our selection, with **the material** we appraise in order to make a selection for preservation. Increasingly, confronted with digital entities, archivists feel that they have first to identify what is the record and, when unable to do so, to determine, on the basis of the creators

functions and activities, whether a record should exist, and if so, to help the creator to redesign the system to enable it to create records that can be preserved and serve a memorial or an evidential function. Now, is this an appraisal activity? I do not think so.

Appraisal is an activity conducted on existing records in relation to values defined by an archival policy. However, this new activity does occur in the course of the assessment of the records for selection and preservation purpose. The issue is quite difficult when, for reasons primarily of social and professional responsibility, but also of transparency and accountability, records must be created that contain a certain type of data and must be kept for the same use for which they are created, probably for a very long time. In such case, appraisal, that is, the assessment of the value of specific data sets, is conducive to the definition of the form of the records that should contain them and of the digital presentation that will allow for their long term use, accessibility and preservation. Although appraisal ends up serving a creation purpose and is followed by the actual creation of new documentary forms, it does not affect the impartial, involuntary nature of the resulting records, as the data sets already exist and the records that come to contain them are used in the usual and ordinary course of business by the creators for their own purposes...they are not generated for research purpose.

These points are easily demonstrated by two case studies carried out by InterPARES 2, the already mentioned VanMap, and the Alsace-Moselle Land Registry.⁶

Appraisal, differently from selection, is entirely conditioned by context, and requires two parties, the creator and the preserver, who must have a very clear, transparent relationship. Precisely because of the fragility and volatility of digital records, trust is paramount and must permeate every activity carried out on them.

In light of the above, can we state that there is continuity in the role of the archivist, or are we witnessing a fundamental transformation? The answer depends on the time span we are looking at. If we only consider the past century, I would say that we are confronted with a radical change. The primary role of future archivists is going to be that of *trusted custodians of society's records*. In an archival environment that is becoming predominantly digital, the archivist will be responsible for taking physical and legal custody of, and preserving (i.e., protecting and ensuring continuous access to) the creators' records as soon as they are no longer needed for the usual and ordinary course of activity. Whether an employee of an archival institution or a professional responsible for an in-house unit, the role of the archivist should be that of an officer who guarantees the authenticity of the record in his/her care. To be considered a trusted custodian, the archivist will have to

- act as a neutral third party, i.e., demonstrate that it has no stake in the content of the records and no reason to alter records under its custody, and that it will not allow anybody to alter the records either accidentally or on purpose,
- be equipped with the knowledge and skills necessary to fulfil its responsibilities, which should be acquired through formal education, and
- establish a trusted preservation system that is capable of ensuring that accurate and authentic copies of the creator's records are acquired and preserved.

⁶ See http://www.interpares.org/display_file.cfm?doc=ip2_cs24_final_report.pdf and http://www.interpares.org/ip2/ip2_case_studies.cfm?study=17. The two cases are also discussed in the power point presentation in the CD ROM.

This role of trusted custodian is very new to a world used to regard the archivist as a historical researcher or a facilitator of access, but, if you go further back in history, the *quaestores* of the Roman republic had exactly such function, because the records they were responsible for were written on wax, a very volatile medium. These records were brought to the archives through a subterranean tunnel as soon as they had exhausted their active use within the administrative unit generating them so that nobody would either by accident or maliciously delete or modify them. If a citizen needed to consult them, the *questores* would ask scribes to prepare copies that they would then authenticate. One might also look at the archives of medieval monasteries, where monks would preserve documents by making authentic copies (*cartularia*), *regesta*, indexes, etc., which would make full faith and credit in case of loss or deterioration of the originals. Also Chinese archivists, at the end of each dynasty, used to make copies and summaries of the most important records, which were considered the most trusted source for the history of the dynasty. Thus, yes, we do have continuity over a couple of millennia but a definitive break with the image of the archivist we are most familiar with.

At the conclusion of its second phase InterPARES has reaffirmed a few points of principle that it 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, 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; and, fourth, such solutions are always specific, thus action plans must be developed by each archival program or institution in relation to the records it acquires in the context of the creators for which it is responsible.

To help programs and institutions to meet the latter challenge, InterPARES has initiated in 2007 its third phase. In fact, a study of the effectiveness of workshop and seminar experiences for increasing archivists' skills in digital preservation and their ability to implement these skills in their repositories has shown that very few participants were able to implement the skills once they returned to their work environments.⁷ Moreover, feedback on the outcomes of the two phases of InterPARES from archivists working in institutions smaller than national archives has consistently shown concern about their downward-scalability and their relevance to small and medium sized organizations.

Thus, the goal of InterPARES 3 is to enable small and medium-sized public and private archival organizations and programs (units within records creating organizations)— which are responsible for the digital records resulting from government, business, research, art and entertainment, social and/or community activities—to preserve over the long term authentic records that satisfy the requirements of their stakeholders and society's needs for an adequate record of its past.

The expected products are:

⁷ Wendy M. Duff, M., Amy Marshall, Carrie Limkilde and Marlene van Ballegooie (2006). "Digital Preservation Education: Educating or Networking?" *The American Archivist* 69(1): 188-212.

1. Policies, strategies and procedures for small archival organizations or programs, and guidelines for the records creators whose records fall under their responsibility.
2. Action plans for the specific case studies carried out in the course of the Project.
3. Criteria to determine “most-at-risk” materials e.g., checklist of age (date created, date last accessed), physical carrier, operating system, software used, equipment required and its availability, etc.
4. Guidelines for addressing digital preservation requirements that apply to specific types of records, but not to other materials.
5. Evaluation models for assessing the degree of success, if any, of the chosen preservation action.
6. Cost-benefit models for various types of archival organizations or programs and for various kinds of records and/or systems.
7. Ethical models that identify and make explicit the consequences for individuals and society of types of preservation measures or lack thereof
8. Training and education modules for preservers, professional associations and university programs; and awareness and education modules for non-archivists, such as IT professionals, vendors and service providers; human resources and financial managers; doctors, communities of practice, members of the general public, etc.; and a strategy for delivering them.
9. Position papers directed to key regulating, auditing and policy-making bodies, advocating the vital need of embedding planned digital preservation in the requirements they issue for the activities they regulate, audit or control.⁸

The project has started a year ago and we already have some important findings, mostly alarming, but useful nonetheless. Among the most positive outcomes of the initial case studies is the fact that the InterPARES guidelines for records creators and records preservers, as well as the policy framework, are very adaptable instruments that can be profitably used by all kinds of small and low resources organizations. Among the least positive findings is the enormous impact that organizational culture has over the applicability of standards and good practices. The amount of resistance encountered by any proposition of change is equal to the desire that organizations have of implementing change to solve the problems presented by digital records. This is why the most important component of InterPARES 3 is the educational one. Thus, I will conclude with an appeal to you all: spread the word as much and as far as you can, because at this point doing nothing about digital records is equivalent not to staying still while technology progresses, but to running backwards.

⁸ http://www.interpares.org/ip3/ip3_index.cfm