

The concept of record in experiential, interactive and dynamic environments: can the InterPARES project address the ultimate archival challenge?

Luciana Duranti

Chair and Professor, Archival Studies
Director, InterPARES Project <www.interpares.org>
School of Library, Archival & Information Studies
Suite 301 - 6190 Agronomy Road
The University of British Columbia
Vancouver, B.C. V6T 1Z3
Canada
Tel: +1-604-822-2587
E-mail: luciana@interchange.ubc.ca

RESUMO

This paper discusses the concept of electronic record as articulated and used in the context of the InterPARES Project, a multinational and multidisciplinary research project that aims at developing the theoretical and methodological knowledge essential to the long-term preservation of authentic records created and/or maintained in digital form. This knowledge should provide the basis from which to formulate model policies, strategies and standards capable of ensuring the longevity of such material and the ability of its users to trust its authenticity. InterPARES has developed in two phases, the first of which was concerned with electronic records created and/or maintained in databases and document management systems, and the second with electronic records existing in experiential, interactive and dynamic digital systems. The paper describes the characteristics, elements, attributes and components of electronic records and, doing so, it shows how the concept of record in the electronic environment is at the same time much more precise than in the traditional one, and in constant evolution.

PALAVRAS-CHAVE: Guia, instruções, actas de congresso

It is a common saying that if a given entity looks like A, smells like A, and behaves like A, it is A. When it comes to records, this saying used to be absolutely true. With digital entities, it is no longer. What the InterPARES research project has found, one case study after another, is that when a digital entity looks like a record, behaves like a record, is treated by the creator like a record, and even smells (yes, there is software capable of conveying that dusty mushy smell) like a record, may or may not be a record. And, conversely, if a digital entity does not look like a record, does not behave like a record, and is not treated like a record, it might very well be a record. But I am getting ahead of myself. I should first introduce briefly the InterPARES Project.

The InterPARES (International research on Permanent Authentic Records in Electronic Systems) Project is a research endeavour that aims at developing the theoretical and methodological knowledge essential to the long-term preservation of authentic records created and/or maintained in digital form. This knowledge should provide the basis from which to formulate model policies, strategies and standards capable of ensuring the longevity of such material and the ability of its users to trust its authenticity. InterPARES has developed in two phases. It started out to deal with records mandated for accountability and administrative needs. In most countries, such records are the majority of those selected for permanent preservation, and constitute a high priority for both the public and the private sector. When in electronic form, they are usually created in databases and document management systems. The authenticity of these records on traditional media has been a concern of most juridical systems, which have explicitly stated requirements for their authenticity that could be used as a starting point for developing new requirements for their electronic counterparts. The creation, maintenance and use of these records are highly controlled, thus the first phase of InterPARES was able to focus on the preservation of the authenticity of records that are no longer needed by the creating body to fulfill its own mission or purposes, and issued authenticity requirements, and methods of appraisal and preservation. However, by the time this phase was concluded, the electronic records produced in the normal course of affairs had become much more complex. Thus, a second phase of research began, InterPARES 2, the goal of which is to ensure that the portion of society's recorded memory that is digitally produced in interactive, dynamic and experiential systems in the course, and as a byproduct, of artistic, scientific and electronic government activities can be created in accurate and reliable form, and maintained and preserved in authentic form, both in the short and the long term, for the use of those who created it and of society at large, regardless of

digital technology obsolescence and media fragility. The research objectives of InterPARES 2 are:

- to develop an understanding of interactive, dynamic and experiential systems and of the records produced and maintained in them, of their process of creation, and of their present and potential use in the artistic, scientific and government sectors;
- to formulate methods for ensuring that these records are generated and maintained by the creator in such a way that they can be trusted as to their content (that is, are accurate and reliable) and as records (that is, are authentic);
- to formulate methods for selecting among them those that have to be kept after they are no longer needed by the creator in the ordinary course of activity because of their legal, administrative, social or cultural value;
- to develop methods and strategies for keeping the records selected for continuing preservation in authentic form over the long term;
- to develop processes for analyzing and criteria for evaluating advanced technologies for the implementation of the methods listed above in ways that respect cultural diversity and pluralism; and
- to identify and/or develop specifications for policy, metadata, and automated tools necessary for the establishment of an electronic infrastructure capable of supporting the creation of accurate and reliable, and the preservation of authentic digital records.

The InterPARES research team determined at the very beginning of the first phase of the project that, before plunging into the study of the material in question, it was necessary to agree on the concept of record and on how it differed from document, information and data. Thus, the team decided to define a record as any document created (i.e., made or received and set aside for further action or reference) by a physical or juridical person in the course of a practical activity as an instrument and by-product of it, thereby choosing the traditional archival concept. The team then proceeded to define document as recorded information, information as a message intended for communication across space or time, and data as the smallest meaningful piece of information. Finally, an electronic record was defined 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 for action 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, the research focused on records born, maintained and used in electronic form. Regardless of the choice of a traditional archival definition for an electronic record, it was essential to determine what the necessary characteristics of such record are. The following were identified: 1) a fixed form, meaning that the binary content must be stored so that it remains complete and unaltered, and its message

can be rendered with the same documentary form it had when first set aside; 2) an unchangeable content; 3) explicit linkages to other records within or outside the digital system, through a classification code or other unique identifier; 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.

Having specified the necessary characteristics of a record, the research team made the fundamental assumption that, regardless of differences in nature, provenance or date, all records are similar enough to make it possible to conceive of one typical, ideal documentary form containing all possible elements of a record. On the basis of this assumption, we hypothesized that, while they may manifest themselves in different ways, the same elements that are present in traditional records exist either explicitly or implicitly in electronic records, and that all electronic records share the same elements. Thus, we created a template, that is, a decomposition of the ideal electronic record into its constituent elements, which defines each element, and explains its purpose. We used it as an instrument for the systematic analysis of the electronic entities contained in several different systems, for the purpose of establishing which ones are records.

The template is composed of four sections: documentary form, annotations, context, and medium. The documentary form includes, among the internal elements, the names of the persons concurring to the creation of the record, the chronological date, the indication and description of the action or matter, the attestation and a statement of validation, and, among the external elements, overall presentation features (e.g. text, image, sound, graphic), specific presentation features (e.g. layouts, hyperlinks, colors, sample rate of sound files, resolution of image files, scales of maps), electronic signatures and seals (e.g. digital signature), digital time stamps, and special signs (e.g. digital watermarks, organization crest, personal logo).

The annotations fall into three fundamental groups: 1) additions made to the record after its creation as part of its transmission (e.g. priority of transmission, date of compilation and date of transmission in an e-mail record, the indication of attachments), 2) additions made to the record in the course of handling the business matter in which the record participates (e.g. date and time of receipt, action taken, name of handling office), and 3) additions made to the record in the course of managing it as a record (e.g. filing date, class code, registration number). The categorization of the contexts of the record and the list of what would reveal them correspond to an hierarchy of frameworks that goes from the general to the specific: 1) juridical-administrative context (e.g. laws and regulations), 2) provenancial context (e.g. organizational charts, annual reports, tables of users in a database), 3) procedural context (e.g. workflow rules, codes of administrative procedure), 4) documentary context (e.g. classification schemes, records inventories, indexes,

registers), and 5) technological context (e.g. hardware, software, system models, system administration).

The medium was difficult to place within the template, because, although it is still necessary for an electronic record to exist, it is no longer inextricably linked with the message, does not store the record as such, but a bit-stream—because the record, to be complete, needs the software that reads it, and its choice by the record-maker or keeper can be completely arbitrary or based on reasons related to preservation rather than to the function of the record. In addition, the medium is not a relevant factor in assessing a record's authenticity, at least from the perspectives of the creator and of the record preserver. This was confirmed by the case studies, by the end of which the research team was convinced that, with electronic records, the medium should be considered part of the technological context.

The analysis conducted using the template indicated that only a half of the examined systems contained records (twelve out of twenty-two), primarily because the entities identified within the other half did not appear to possess either a fixed documentary form or a stable content. When systems did contain records, these could rarely be compared with the model represented by the template, because, although they were able to achieve their purposes, they were not good records. For example, in most systems, there was no explicit manifestation of the relationship among the records participating in the same affair, and, although it was easy to identify the business processes supported by the system, it was not always possible to determine how the records participated in or supported specific actions. In addition, it was often difficult to determine the significance of the presence or absence of given elements of documentary form or of annotations.

More importantly, the case studies showed that, with digital records, a key component is provided by the records attributes, that is, the defining characteristics of each given record or of a record element in it. A *record element* is a constituent part of the record's documentary form and may be either extrinsic, like a seal, or intrinsic, like the salutation. An *attribute* may manifest itself in one or more elements of a record's documentary 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. Every record has attributes that manifest themselves in formal elements, that is on the face of the record, and attributes that are implicit in other components of the record, such as the name of the medium, but in digital records they are also formally expressed outside the documentary form. However, they are mostly transparent to the user, and manifest themselves as metadata included in either a record profile, a topic map, or other digital entity linked to the record. Attributes, or metadata, if you wish, are important to identify any digital record, but they become the primary means for the identification of digital records that do not have a stable form, or fixity. This will become clearer later on.

As if the distinction between record elements and attributes were not confusing enough, with electronic records, record elements must also be differentiated from the record 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 record, and that requires specific methods for preservation. For example, an e-mail containing text, a picture and a digital signature has at least four digital components that are stored in different part of the system, although they are linked among themselves, and require different protection measures: the header, the text of the message, the picture and the digital signature. In contrast, a report with textual attachments may consist of only one digital component. In other words, a digital components is a unit of storage, but one that needs to be identified when the concept of digital record is dissected.

Finally, it needs to be pointed out that the relation between a digital record and a computer file can be one-to-one, one-to-many, many-to-one, or many to many, thus we should never use the terms record and file interchangeably; that the same presentation of a record can be created by a variety of digital presentations and, vice-versa, from one digital presentation a variety of record presentations can derive; and that it is possible to change the way in which a record is contained in a computer file without changing the record, thus the name of a record form does not necessarily indicate what record we are dealing with.

The concept of digital record presented so far, with all its characteristics, parts, elements, attributes and components, works quite well with databases and document management systems. However, it may appear problematic when applied to the records examined by InterPARES 2, the most salient characteristic of which is, as mentioned earlier, the lack of a stable form and content. They are experiential, interactive and dynamic records.

Experiential records are electronic objects the essence of which goes beyond the bits that constitute the object to incorporate the behavior of the rendering system, or at least the interaction between the object and the rendering system. Defining the characteristics, elements, attributes and components of such objects is much more complex than with raw data or more traditional electronic records, because it is dependent not only on the object per se, but on the environment in which the object is experienced. Examples of experiential digital objects range from audio and moving images embedded in a web page to virtual reality systems.

Interactive records are records made and maintained in interactive systems, where each user's entry causes a response from or an action by the system. To determine the boundaries of such records (i.e., where one record ends and another begins), when they can be considered finished rather "in progress", when they are complete rather than partial or incomplete, etc., one needs to ascertain 1) how user input affects the creation and form

of each record (as is the case with much on-line commerce); and 2) if and when the interactive system and its inherent functionality are to be regarded as meaningful parts of the record. Examples of interactive systems range from web pages delivering government services online to musical performances based on human-computer interaction to commercial video games.

Dynamic records are documents whose content is dependent upon data that vary continuously and are held in several databases and spreadsheets. Examples range from simple web pages with embedded links to complex systems where information is stored and updated to be shared via wireless transmission by multiple mobile users in diverse ways. The increasing reliance on such documents by individuals and institutions will necessitate understanding how the information they contain is captured and saved.

Whether experiential, interactive, and dynamic digital objects are indeed records in the first place primarily depends on their relationship to the activity of their creator. Ironically, the ease with which their form and content can be manipulated has given those who generate them a new reason for keeping them: 'repurposing'. Records creators often obscure the meaning and cultural value of their records by treating their content merely as digital data to be manipulated to generate new records, decontextualizing them from the activity by which they were produced. In fact, we are increasingly looking at digital dynamic "records" whose "recordness" is constantly destroyed in one context and reconstituted in a new context. The potentially wide dissemination of repurposed records threatens the continuing existence of the materials subject to this treatment and it is another issue to wrestle with.

In light of these new types of records, it is probably necessary to revisit the concept of record itself, so that both the identification and the protection of experiential, interactive and dynamic documentary information will be possible. It appears that the concept of record developed in the first phase of InterPARES, although very useful when applied to document management systems and databases, is limited in its capacity to examine electronic systems containing a variety of very complex entities. This of course shows how not always that which is known is helpful to understand that which is unknown. Thus, to make our methodology of analysis more useful, the structure of our template, representing the ideal, all encompassing digital record, needs to be more flexible, and the identification and interpretation of its components, elements, attributes, etc. should be more nuanced. In addition, we should consider the possibility of trading the characteristics of completeness, stability and fixity with the ability of the system to track and preserve any change to the records. That is, we should consider the possibility of attributing completeness, stability and fixity to the record metadata, rather than to the record form and content. And perhaps we may look at the record as existing in one of two modes, as an entity *in fieri*, in becoming, when its process of creation is in

course (even if such process is ongoing), and as a fixed entity at any given time the record is used. There is no doubt that knowledge and strategies must be developed that are beneficial for both the creators and preservers of these complex new records. One way of doing so is to keep in mind that many of the issues surrounding the management of electronic records in the arts and sciences are becoming relevant to government archives, because administrative bodies are increasingly employing complex multimedia systems in the creation of their records. In Canada, for example, the Government On-line initiative has mandated that most transactions between the government and its citizens be possible on the Internet in an interactive mode by 2006. This raises considerable questions for the creation and management of the electronic records generated by such interaction, in part because the making of the record will no longer be the sole responsibility of the body having control of the electronic system (in this case, the government), but also of the user. Additional questions are raised by the double public and private nature of these records that would be shared by private and public persons on-line, rather than existing as distinct entities in the archives of each person participating in the transaction. Further, when the terms and conditions that govern the recorded transactions between government and citizens are articulated on web pages, the functionality of those pages may have to be inextricably linked to the resulting records for purposes of accountability.

Questions whose answers generate more questions. Is the record in the interactive environment a mere surrogate? In other words, should we regard metadata records of experiential, interactive and dynamic digital entities as having the same function and authority as the registers of medieval times? Do the new complex records have any meaning and authority when extracted from their own digital environment? Where does the responsibility of the archivist begin and where does it end in influencing the form and process of creation of the records that will eventually fall under his jurisdiction? We seem to be constantly trying to catch up when it comes to the challenges presented by new information technologies, but I am convinced that, now as well as in the foreseeable future, the ultimate challenge of the archivist is to identify the object of its care, and to recognize and isolate it in the immense flux of unending information. And this is probably the area in which the continuing efforts of InterPARES might be the most fruitful: perhaps, when all the findings of InterPARES 2 will be in and out, archivists will be again certain that if it looks like a record, sounds like a record, smells like a record, behaves like a record, and is treated like a record, it cannot be anything else than a record.