

Reflections on InterPARES



Archivaria

The InterPARES 2 Project (2002–2007): An Overview

LUCIANA DURANTI

RÉSUMÉ La deuxième phase du projet InterPARES s'appuie sur les conclusions de la première phase (1999–2001) afin d'aborder le défi de la préservation permanente de documents numériques fiables, exacts et authentiques, créés et maintenus dans des systèmes interactifs et dynamiques au cours de toutes sortes d'activités humaines. Ce survol décrit le but et les méthodes de ce projet, et il dresse les grandes lignes des conclusions et des produits afin de fournir un cadre de travail pour les articles publiés dans *Archivaria*, dans la section intitulée « Réflexions sur InterPARES ».

ABSTRACT The second phase of the InterPARES Project built upon the findings of the first phase (1999–2001) to address the challenge of the permanent preservation of reliable, accurate, and authentic digital records created and maintained in interactive and dynamic systems in the course of all kinds of human activities. This overview describes the goal and methods of the project, and outlines findings and products to provide a framework for the articles published in the *Archivaria* section entitled, "Reflections on InterPARES."

The first phase of the InterPARES (International Research on Permanent Authentic Records in Electronic Systems) Project (1999–2001) primarily addressed the permanent preservation of textual records that were born digital in the course of administrative and legal activities in databases and document management systems, and had reached an inactive status. In relation to them, it developed the concepts of digital record and authenticity; authenticity requirements for those who generate and keep records and for those who preserve them (e.g., metadata for identity and integrity, access privileges, etc.); methodologies of appraisal and preservation from the preserver point of view; a series of analytical instruments for studying new types of digital documents and developing new requirements and methods as needed; and a framework for the development of policies, strategies, and standards related to the

proper creation, maintenance, and preservation of digital records that can be proven authentic over time.¹

InterPARES 1 research found that most systems that should contain records do not, because the entities in them lack fixed form and stable content. Moreover, the systems that do contain records have bad ones, primarily because of the absence of identifiable contexts and relationships. Further, it found that inactive records that are no longer kept in active systems often cannot be preserved because either they were not created and/or maintained in preservable formats or they are obsolete; therefore, the preservation of authentic electronic records must be a continuous process that begins with records creation, must be predicated on the concepts of trusted record-making, record-keeping and record-preservation systems, and on the role of the preserver as a trusted custodian. It must incorporate records appraisal and archival description as integral components and necessary instruments for reaching its purposes.

InterPARES 1 researchers also concluded that the classic concept of record had limited their capacity to understand electronic systems containing a variety of complex entities that do not correspond to it, and that theory, which decontextualizes the record, is not always useful to deal with the variety and complexity of digital systems used in the course of human activities. As a result, a complementary inductive approach was necessary. Thus, we developed a new research project, InterPARES 2 (2002–2006), with four major components: 1) it examined all kinds of digital entities in complex systems and subsequently, on the basis of its findings, developed a concept of record reflecting the observed realities; 2) it was concerned with the entire life cycle of the record, to be represented in a chain-of-preservation model that began with the design of the system and addressed accuracy and reliability in addition to authenticity; 3) it studied the digital entities created in the course of activities that have used complex systems well before government and business, that is, e-art and e-science; and 4) it used the concepts and methodology of all the fields touched by our study. Indeed, the goal of InterPARES 2 was to ensure that the portion of society's recorded memory digitally produced in dynamic, experiential, and interactive systems in the course of artistic, scientific, and e-government activities can be created in accurate and reliable form, and maintained and preserved in authentic form, both in the short and long

1 Luciana Duranti, ed., *The Long-Term Preservation of Authentic Electronic Records: The Findings of the InterPARES Project* (San Miniato, 2005), also online at <http://www.interpares.org/book/index.cfm> (accessed 15 August 2005).

term, for the use of those who created it and of society at large, regardless of digital technology obsolescence and media fragility.²

The integration of artistic and scientific activities into the core research endeavour of the project required the participation of researchers who were experts in those specific fields and in the design of complex digital technology: two thirds of InterPARES 2 researchers were not from the records and information professions, but scholars of music theory, composition, and performance; film theory, production, and description; dance and theatre theory; photography; media science; a variety of hard and social sciences (such as chemistry, astronomy, geography, linguistics, and archaeology); jurisprudence; computer science; and engineering. The reason for this composition of the research team went beyond the fact that, in order to preserve records, we need to understand the nature of the activities generating them and their function and use in the context of those activities.

By the end of InterPARES 1, it was clear that the solution to the problems presented by digital preservation could not be found in the context of a strictly disciplinary approach, but needed the support of a variety of disciplines whose body of knowledge had focused on individual aspects of those problems. This understanding required an *interdisciplinary* approach, that is, a way of conducting research that would integrate several methodologies, concepts, principles, and techniques from a variety of fields as needed, elaborate and develop its results, and produce new knowledge consistent with that of each separate field. However, in the course of the research, we went much beyond interdisciplinarity, experimenting with multidisciplinary and transdisciplinarity. The *multidisciplinary* process, which we adopted for gaining an understanding of the concepts of reliability, accuracy, and authenticity across disciplines, examines the same problem in the context of each separate discipline and aims at solving it within each discipline, without any integration with the theory or method of another discipline. Then the results are compared and the best solutions for the specific purpose adopted.

In comparison with interdisciplinarity and multidisciplinary, *transdisciplinarity* is multi-referential and multi-dimensional. Whereas interdisciplinarity involves the transfer of one or more methods or ideas from one discipline to another, and multidisciplinary involves the analysis of the same object by many disciplines, transdisciplinarity, as the prefix “trans” indicates, involves thinking at the same time within, across and outside each discipline, and

2 See <http://www.interpares.org/> (accessed 15 August 2007). Both phases of the InterPARES Project have been financially supported by the Social Sciences and Humanities Research Council of Canada (SSHRC), and by several other agencies who have funded our international partners, such as the US National Science Foundation (NSF) and the US National Historical Publication and Records Commission (NHPRC), and the Associazione Nazionale Archivistica Italiana (ANAI).

beyond all disciplines. Its purpose is to gain an understanding of present reality, one imperative of which is the unity of knowledge.

Rigor, *openness*, and *tolerance* are the fundamental characteristics of the transdisciplinary attitude and vision. *Rigor* in argument, taking into account all existing data, is the best defence against possible distortions. *Openness* involves an acceptance of the unknown, the unexpected and the unforeseeable. *Tolerance* implies acknowledging the right to ideas and truths opposed to our own.³

Indeed, transdisciplinarity is the most creative of all research perspectives, because it does not deny disciplinarity while opening the option of rejecting all disciplines; it is therefore the most useful approach when one is confronted with issues that appear, but have yet to be proven, to be entirely new.⁴ We have used this approach particularly in the determination of what is a record in each of the case studies carried out in the course of the project.⁵

Regardless of the number of disciplines involved with the project, the ultimate goal of InterPARES 2 was archival in nature, in that it was concerned with the development of a trusted record-making and record-keeping system⁶ and of a preservation system capable of ensuring the authenticity of the records under examination over the long term. Thus, the work carried out throughout the project in the various disciplinary areas had to be constantly

- 3 See the International Center for Transdisciplinary Research (CIRET), *Charter of Transdisciplinarity*, <http://nicol.club.fr/ciret/english/charten.htm> (accessed 15 August 2007). Every transdisciplinary project is by definition also disciplinary, interdisciplinary, and multidisciplinary.
- 4 A project that has used this approach is Simon Fraser University's Transnet, the Transdisciplinary Network for Performance and Technology, <http://www.sfu.ca/transnet/> (accessed 15 August 2007). A typical interdisciplinary project is ERPANET (Electronic Resource Preservation and Access Network), <http://www.erpanet.org/> (accessed 15 August 2007). Most of the other international collaborations have been multidisciplinary.
- 5 See Luciana Duranti, "Pour une diplomatie des documents électroniques," *Bibliothèque de l'École des chartes*, special issue on "Exportations de la diplomatie" (Paris, 2004), pp. 105–25; Luciana Duranti, "The Long-term Preservation of Accurate and Authentic Digital Data: The InterPARES Project," *Data Science Journal* 4 (2005), pp. 106–18, http://www.jstage.jst.go.jp/articles/dsj/4/0/4_106/_article (accessed 23 August 2007); Luciana Duranti, "La questione fondamentale: in quali entità digitali si concreta la memoria del futuro?," *Archivi & Computer* 1 (2006), pp. 24–31; Luciana Duranti and John Roeder, "La conservazione a lungo termine dell'autenticità di composizioni digitali interattive: la ricerca di InterPARES 2," *Archivi & Computer* (2006), pp. 35–42; Luciana Duranti and Kenneth Thibodeau, "The Concept of Record in Interactive, Experiential and Dynamic Environments: the View of InterPARES," *Archival Science*, vol. 6, no. 1 (March 2006), pp. 13–68, <http://dx.doi.org/10.1007/s10502-006-9021-7>.
- 6 A trusted record-making and record-keeping system comprises the whole of the rules that control the creation, maintenance and use of the records of the creator, and that provide a circumstantial probability of the accuracy, reliability, and authenticity of the records within the system.

translated into archival terms and linked to archival concepts, which are the foundation upon which the systems intended to protect the records should be designed. However, upon completion of the research, the archival systems need to be made accessible and comprehensible to records creators, organizations and institutions, and disciplinary researchers. In other words, the research outcomes must be translated back into the language and concepts of each discipline that needs to make use of them. This *transferability* was ensured by an initial effort to clarify the key archival concepts at the core of the InterPARES 2 research, so that each discipline could identify the corresponding entities within its own body of knowledge, and then by building a Terminology Database, containing a Glossary with the InterPARES definitions of the terms used in the course of the project, a Dictionary with the definitions of the same terms in other disciplines, and Ontologies that illustrate the relationship among related terms.⁷

While InterPARES 1 had its epistemological roots in the humanities, specifically in diplomatics and archival science, InterPARES 2, although planning as one part of its research to test some of the outcomes of InterPARES 1 in a range of applied settings, espoused no epistemological perspective or intellectual definitions *a priori*. Instead, researchers in each research unit identified the perspective(s), research design, and methods that they believed to be most appropriate to their investigation, maintaining therefore a spirit of *open inquiry*. The reason for this openness is that InterPARES 2 was conceived to work as a “layered knowledge” environment, in the sense that some of the research work built upon knowledge developed in the course of the UBC Project⁸ and InterPARES 1; some took the knowledge of similar issues developed in other areas of endeavour and brought it to bear on records creation and preservation; some reconciled knowledge about records and their attributes, elements, characteristics, behaviour and qualities existing in various disciplines, and developed it for archival purposes; and some explored new issues and studied entities never examined before, and therefore developed entirely new knowledge. This layered knowledge environment required a *multi-method research design*. Thus, each research activity was carried out using the methodology and the tools that the dedicated investigating team

7 InterPARES 2 Terminology Database (2007), http://www.interpares.org/ip2/ip2_terminology_db.cfm (accessed 15 August 2007).

8 The UBC-MAS project “Preservation of the Integrity of Electronic Records,” was carried out from 1994 to 1997. See Luciana Duranti, Terry Eastwood, and Heather MacNeil, *Preservation of the Integrity of Electronic Records* (Dordrecht, 2002); and Luciana Duranti and Heather MacNeil, “The Preservation of the Integrity of Electronic Records: an Overview of the UBC-MAS Research Project,” *Archivaria* 42 (Spring 1997), pp. 46–67, <http://journals.sfu.ca/archivar/index.php/archivaria/article/view/12153/13158> (accessed 15 August 2007).

considered the most appropriate. The methods used included surveys, case studies, general studies, modeling, prototyping, diplomatic and archival analysis, and textual analysis.

The research was guided by the research questions that were developed for the project proposal. These questions were reformulated in ways understandable to the researchers of all participating disciplines, who, in preparing their case- and general-studies tools and framing their inquiries, were also guided by additional research questions jointly developed by the international team in light of issues raised at the initial meetings.

The objects of InterPARES 2 inquiry included: dynamic entities, depending for their content upon data extracted from a variety of systems which may have variable instantiations (e.g., the VanMap discussed in the article by Glenn Dingwall, Richard Marciano, Reagan Moore, and Evelyn Peters McLellan appearing elsewhere in this issue); experiential entities, whose essence goes beyond the bits constituting it to incorporate the behaviour of the rendering system and the effects of subjective user interactions (e.g., the Cybercartographic Atlas of Antarctica discussed in the article by Tracey P. Lauriault, Barbara L. Craig, D.R. Fraser Taylor, and Peter L. Pulsifer appearing elsewhere in this issue); interactive entities, to which each user intervention or input from another system causes a change of content and/or form (e.g., the Alsace-Moselle Land Registry, discussed in Duranti-Thibodeau⁹); live, active materials, followed from creation to preservation (e.g., the Electronic Engineering and Manufacturing Records assessed in Case Study Nineteen by Kenneth Hawkins); obsolete inaccessible materials (e.g., the Obsessed Again music piece¹⁰); categories of records creators, like composers, photographers, and archaeologists¹¹; digital file

9 See Duranti and Thibodeau, "The Concept of Record," p. 30, and "Case Study Characterization, CS 18, Computerization of Alsace-Moselle's Land Registry," http://www.interpares.org/display_file.cfm?doc=ip2_alsace_characterization.pdf (accessed 15 August 2007). All links to case studies can also be found at http://www.interpares.org/ip2/ip2_case_studies.cfm (accessed 15 August 2007).

10 See "Completed Report for the Obsessed Again Case Study," [http://www.interpares.org/display_file.cfm?doc=ip2_obsessed_again\(complete\).pdf](http://www.interpares.org/display_file.cfm?doc=ip2_obsessed_again(complete).pdf) (accessed 15 August 2007).

11 See Michael Longton, "Completed Report for the Survey of Record-keeping Practices of Composers General Study," February 2004, [http://www.interpares.org/display_file.cfm?doc=ip2_composers_survey\(complete\).pdf](http://www.interpares.org/display_file.cfm?doc=ip2_composers_survey(complete).pdf); Jessica Bushey, "Survey of Recordkeeping Practices of Photographers using Digital Technology Final Report for Participants," 28 August 2005, http://www.interpares.org/display_file.cfm?doc=ip2_digital_photo_recordkeeping_report.pdf; Randy Preston, "Survey of the Digital Recordkeeping Practices of GIS Archaeologists Worldwide: Results of a Web-based Survey," 15 February 2007, [http://www.interpares.org/display_file.cfm?doc=ip2_gis_arch\(complete\).pdf](http://www.interpares.org/display_file.cfm?doc=ip2_gis_arch(complete).pdf).

formats¹²; and several approaches taken by both creators and preservers¹³ who have concerned themselves with accuracy, reliability, and authenticity.

The formal findings and products of the InterPARES 2 Project are multiple and varied. They will be presented in a book, including not only eight reports by the various research units, but also specific papers on case and general studies, that are published electronically on the project's website and, in print (with enclosed DVD), by Associazione Nazionale Archivistica Italiana (ANAI). The products of InterPARES 2 that can be used as autonomous resources include: a *Framework of Principles Guiding the Development of Policies* for records creating and preserving organizations; *Guidelines for Making and Maintaining Digital Records* for individuals and small communities of practice; *Guidelines for Digital Preservation* for archival institutions; a *Metadata Registry* for the registration and analysis of metadata schemas; a *Chain of Preservation Model*; principles and criteria for adoption of *File Formats, Wrappers, and Encoding*; and a *Terminology Database*. All these resources can be freely downloaded from the InterPARES 2 website, and used as needed without any concern for copyright. Their purpose is to disseminate as widely as possible the knowledge developed by InterPARES researchers.

Differently from the book and the resources mentioned above, the articles that appear in this section of *Archivaria* go beyond the formal findings and the products of InterPARES 2 to discuss ideas and thoughts that either individual scholars or groups of researchers entertain about the specific research activity in which they have participated. They raise issues or present challenges that have not been resolved within the project, either to their satisfaction or not at all; questions that have not been addressed because they were outside the scope of the project, but considered significant in their specific context; or matters whose discussion did not find a large enough place within the InterPARES final documents, but deserve special attention. Also differently from the material that appears in the book and in the resources, the content of these articles and its form of expression have not been scrutinized and approved by the entire InterPARES team as the official, shared result of the research project, but is the product of individual or collaborative authorship as attributed in each article, and is indebted to the project as a whole only in the measure identified by the author(s) in the credits expressly given in the foot-

12 See Evelyn Peters McLellan, "Selecting Digital File Formats for Long-Term Preservation: InterPARES 2 General Study 11 Final Report," [http://www.interpares.org/display_file.cfm?doc=ip2_file_formats\(complete\).pdf](http://www.interpares.org/display_file.cfm?doc=ip2_file_formats(complete).pdf) (accessed 15 August 2007).

13 For example, see Regan W. Moore, "Building Preservation Environments with Data Grid Technology," *American Archivist*, vol. 69, no. 1 (July 2006), pp. 139–58.

notes. That said, one example of the way these articles relate to the core InterPARES 2 knowledge formally issued in the official research products is provided by the way they deal with the concept of record in the environments they examine.

After analyzing all the data provided by the numerous case studies and general studies, InterPARES 2 concluded that, in order to be defined as records, digital entities must be affixed to a medium and have stable content and fixed form, as well as explicit linkages to other records inside or outside the digital system (i.e., an archival bond), five necessary persons involved in its creation (i.e., author, writer, originator, addressee, and creator), an action in which they participate or which they support, and five necessary contexts of creation (i.e., juridical-administrative, provenancial, procedural, documentary, and technological). In light of the fact that we were dealing with records in dynamic and interactive systems and that the material we had examined revealed a clear difference between a manifested record and a stored record,¹⁴ with cases in which one of the two was not present, we further elaborated the concepts of stable content and fixed form. We stated that “stable content” means that the data and the message in the record are unchanged and unchangeable, meaning that data cannot be overwritten, altered, deleted, or added to; and that “fixed form” means either that the binary content of the record is stored so that the message it conveys can be rendered with the same presentation it had on the screen when first saved (different digital presentation), or that the same content can be presented on the screen in several different ways in a limited series of possibilities, resulting in a different view of the same record having stable content and fixed form (different documentary presentations: e.g., statistical data seen as a pie chart, a bar chart, or a table). We then added the concept of “bounded variability” as one satisfying the requirement of stable content and fixed form. Bounded variability is present 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.

14 A manifested record can be defined as 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 recreated 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). A stored record is defined as the digital component(s) used in reproducing one or more than one record and which include the data to be processed in order to reproduce the manifested record and the rules for processing the data, including those enabling variations. Sometimes, the stored record has only an enabling function and does not correspond to any given manifested record.

Regardless of the flexibility offered by the concepts of stable content and fixed form expressed above, the limitations imposed by this specific requirement on the concept of record, and on the consequent method of maintaining and preserving the dynamic and interactive entities¹⁵ that some researchers were studying, created several problems, especially in the scientific environment. While many of the issues presented in that regard by the VanMap case study, for example, could be addressed talking about “records in becoming” and “potential records” until such time when the digital entities are completed, closed, and stabilized by the creator in order to maintain them in a way that satisfies accountability, the issues presented by the Cybercartographic Atlas of Antarctica, in a contrasting example, remained unresolved, with the creator considering its digital entities to be records, whereas, according to the requirements for records described above, they are fluid data.

As I do not wish to spoil the pleasure of *Archivaria* readers, I will end my discussion of this and other issues identified as controversial by the authors of the articles that follow, and will let the readers find out by themselves how the InterPARES team spent hundreds of meeting hours: even when we did not reach an agreement, there is no doubt in my mind that it was a very productive time, as evidenced by this series of articles, and by the fact that it saw scientists, artists, and administrators passionately discuss issues that had only been debated in the arena of archival discourse.

15 Interactive entities present variable content, form or both, and the rules governing the context and form of presentation may be either fixed or variable. They can be static or dynamic. They are static 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 (e.g., web pages, on-line catalogues, and patches enabling performances). They are dynamic if the rules governing the presentation of content and form vary. There are cases where the variation is due to data that change frequently (e.g., the design permits updating, replacement or alterations, allows data collection from users or about user interactions or actions, or uses these data to determine subsequent presentations); cases where the variation is due to data received from external sources and not stored within the system (e.g., VanMap); cases of 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 cases of entities produced by evolutionary computing where the software generating them can change autonomously (e.g., scheduling and modelling of financial markets; edutainment sites). See Duranti and Thibodeau, “The Concept of Record in Interactive, Experiential and Dynamic Environments.”