

# THE LONG-TERM PRESERVATION OF ACCURATE AND AUTHENTIC DIGITAL DATA: THE INTERPARES PROJECT

*L. Duranti*

*School of Library, Archival and Information Studies, The University of British Columbia, Vancouver, British Columbia, Canada  
Email : luciana@interchange.ubc.ca*

## ABSTRACT

*This article presents the InterPARES Project, a multidisciplinary international research initiative aimed at developing the theoretical and methodological knowledge necessary for the long-term preservation of digital entities produced in the course of business or research activity so that their authenticity can be presumed or verified. The methodology, research activities, preliminary findings and projected products are discussed in the context of the issues that the project attempts to address.*

**Keywords :** Electronic records, Digital data, Digital preservation, Accuracy, Authenticity, Metadata, Digital obsolescence

## 1 INTRODUCTION

Ongoing technological change is causing widespread concern around the world regarding the long-term preservation of the material produced or stored using digital technologies. A portion of our society's recorded memory created and preserved digitally has already been compromised, and there are enormous costs associated with recovering electronic entities that have become inaccessible. While the extent to which valuable digital material has been lost or has become retrievable only at great expense has yet to be adequately quantified, it is already apparent that the threat is real and widespread. Moreover, even if we could ensure the preservation of electronic entities and overcome media fragility and technological obsolescence, preserved materials would be of little value unless we can be sure that they are 1) accurate, that is, precise and free of error or distortions, and 2) authentic, which means that their identity and their integrity have not been inadvertently or maliciously compromised, and that they are what they purport to be, immune from corruption and tampering (Council on Library and Information Resources, 2000). For centuries, our presumption of accuracy and authenticity has been premised on the presence or absence of visible formal elements on the documents we were examining and on an uninterrupted line of legitimate custody of such documents. The use of digital technology has not only reconfigured those formal elements, allowed for the bypassing of production controls, and made physical custody an elusive concept, but, first and foremost, it has eliminated the original, that is the first complete instantiation of recorded data communicated either across space (to persons other than the author or originating entity) or time (saved for later access by the author, owner or legitimate successors) (National Research Council, 2000; MacNeil, 2002).

If digital materials will ever be considered as accurate and authentic as those on traditional media, the practices by which they are created, maintained, made accessible and used must be analyzed, and strategies and standards for their preservation must be developed. Many have called for this approach. Among them, representatives of European governments (European Commission, 2000), experts in digital preservation (Dollar, 1999; Levy, 1999), scholars of archival science (MacNeil, 2000; Gilliland-Swetland, 2000), and scholars of records (Ross, 2000). The answer to this call has been InterPARES (International research on Permanent Authentic Records in Electronic Systems), a research endeavor that aims to develop the theoretical and methodological knowledge essential to the permanent preservation of authentic materials generated and/or maintained electronically, and, on the basis of this knowledge, to formulate model policies, strategies and standards capable of ensuring that preservation. At the end of its first phase (InterPARES 1 Project, 1999-2001), InterPARES issued, in addition to methods of selection (Eastwood, 2004a) and preservation, a series of authenticity requirements for materials that, although digital, were very similar to their analog counterparts, especially in that they had a fixed form (Duranti, 2002). Increasingly, however, organizations and individuals have been generating materials of a dynamic, experiential, or interactive nature (Huhtamo, 1999), which will

require different, and perhaps data-type specific, authenticity requirements and selection and preservation strategies.

Dynamic materials depend for their content upon data extracted from databases, which may have variable instantiations (Dale, 1998; Brown, 2000). The challenge they present to those who generate and access them is their lack of fixity, but more serious issues are raised by experiential and interactive objects. Clifford Lynch describes experiential digital objects as objects whose essence goes beyond the bits constituting them to incorporate the behaviour of the rendering system, or at least the interaction between the object and the rendering system. He also maintains that defining the authenticity of such objects is a much more complex problem than with raw data or traditional works, because it is dependent not on the ability to reproduce a copy of the object's original bit-stream, but on the ability to recreate the environment in which that object was experienced, an activity that involves issues of intellectual property, copyright, etc. (Lynch, 2000).

An interactive system is one in which each user's entry causes a response from or an action by the system (Rokeby, 1995; Downes & McMillan, 2000). To generate preservable data in such systems, we need to ascertain a) how user input affects the creation and form of digital data; and b) if and when the interactive system and its inherent functionality need to be preserved for those data to remain meaningful and authentic (Stromer-Galley, 2004).

Whether dynamic, experiential, and interactive digital objects are indeed to be preserved over the long-term depends of course on their relationship to the activity of their creator and on the value that the specific field of the creator and society at large attribute to them (Eastwood, 2004b). Scientific researchers have a long history of creating such objects, and clearly the professionals charged with the preservation of the archives containing them may have to face the concrete challenge of preserving views of dynamic systems, maintaining the functionality of interactive data, and recreating the environment of experiential objects (Moore, 2004). It is important both to know to what extent the requirements, methods and strategies developed by the InterPARES 1 project to preserve authentic electronic material with a fixed form and stable content apply to these new situations, and to develop new ones where they do not. These issues are further compounded when individual creators lack the knowledge and tools to generate data that can be preserved over the long term.

For these reasons, it is necessary to develop an understanding of the new digital objects, not only in the later phases of their life cycle, but from the moment of their creation. In fact, it is probably necessary to revisit the concept of recorded data itself, so that both the identification and the protection of these new types will be possible (Boudrez & Dekeyser, 2003). We have to consider the possibility of substituting the characteristics of stability and fixity of the documents containing the data with the capacity of the system where the data reside to trace and preserve each change each digital object has undergone. And perhaps we may look at each digital entity as existing in one of two modes, as an entity in becoming, when its process of creation is in course (even if such creation is ongoing), and as a finished entity at any given time the data are viewed. There is no doubt that knowledge and strategies must be developed that are beneficial to both the creators and preservers of these complex new materials (Guercio, 2002; Hofman, 2003).

Technological obsolescence, which poses a continual challenge to the accessibility, readability and intelligibility of digital objects, is of even more concern in the context of scientific activities than in that of administrative activities. Several reports of the US National Research Council attest to this. (National Research Council, 1997; 2001a; 2001b). Inadequate information management practices have already precipitated the disappearance of many data sets that depended upon now obsolete software and hardware for their continued existence. This has generated enormous difficulties for scientists concerned with the long-term preservation of the unique and authoritative version of their work, requiring them to devote valuable time and resources to preservation efforts, and engendering an urgent demand for effective and tested strategies (Hodge, 2000; Chen, 2003).

To meet these challenges requires an understanding of the nature of the new electronic objects and their creating processes. Research must be done into their characteristics and development, the requirements for their reliability, accuracy, and verifiable authenticity, and methods and strategies for their selection and preservation. To this end, the international team of researchers formed for InterPARES 1, together with additional researchers with discipline-specific knowledge, decided to initiate a second phase of its research, called InterPARES 2.

## **2 INTERPARES 2: INTELLECTUAL FRAMEWORK**

InterPARES 2 began in 2002 and its completion is scheduled for the end of 2006. Its goal, objectives, structure and methodological principles have been articulated in an intellectual framework on which all co-investigators have agreed.

### **2.1 RESEARCH GOAL**

The goal of InterPARES 2 is to ensure that the portion of society's recorded memory that is digitally produced in interactive, dynamic and experiential systems 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.

### **2.2 RESEARCH OBJECTIVES**

- To develop an understanding of interactive, dynamic and experiential systems and of the materials produced and maintained in them, of their process of creation, and of their present and potential use;
- to formulate methods for ensuring that these digital objects are generated and maintained by the creator in such a way that they can be trusted as to their content (that is, the data in them are accurate and reliable) and as records (that is, the objects are authentic, they are what they purport to be);
- to formulate methods for selecting among them those that have to be kept after they are no longer needed by the creator because of their larger value to research and society at large;
- to develop methods and strategies for keeping the materials 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 creation of an electronic infrastructure capable of supporting the creation of accurate and reliable, and the preservation of authentic digital objects.

### **2.3 GUIDING METHODOLOGICAL PRINCIPLES**

#### **2.3.1 Interdisciplinarity**

The project is interdisciplinary in the measure in which its goal and objectives can only be achieved through the contribution of several disciplines and of all categories of stakeholders: individual creators of digital objects, the information technology sector, the archival and conservation professions, etc. are involved in the formulation and selection of case studies, gathering of empirical evidence, and analysis. Such a mode of research ensures that the project's results will find ready acceptance within the targeted communities. The scholars conducting the research come from the following fields: Archival Science, **Archaeology**, **Astronomy**, **Chemistry**, Computer Engineering, Computer Science, Dance, Diplomats, Film, **Geography**, History, Information studies, Law, Library Science, Linguistics, **Mechanic Engineering**, Media Studies, Music, Performance Art, **Physics**, Photography, **Space Sciences**, and Theatre. The countries actively involved in the project are: Canada, United States, Australia, Belgium, China, France, Ireland, Italy, Japan, Netherlands, Portugal, Singapore, Spain and the United Kingdom. The Advisory Board also includes an archivist from South Africa.

#### **2.3.2 Transferability**

The ultimate goal of the project is archival in nature, in that it is concerned with the development of a trusted system for making and keeping digital entities that are reliable and accurate and of a preservation system that ensures the authenticity of the entities under examination over the long term. This implies that the work carried out throughout the project in the various disciplinary areas must be constantly translated in archival terms and linked to archival concepts, which are the foundation upon which the systems intended to protect the digital entities are 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 need to make use of them. In light of the above, all researchers are committed to learning the key archival concepts that are identified by the archival scholars in the team as constituting the core of the InterPARES 2 research, so that each discipline can identify the corresponding entities within its own body of knowledge. An example of such concepts is the concept of record. Although the traditional definition of the term seems quite generic, its analysis will show that it is loaded with implicit meaning. A record is any document made or received in the course of activity as a means for and by-product of it, and set aside for action or reference. Thus, while not all documents are records, all records are documents first. A document is defined as information which is recorded, that is, affixed to a medium, and this implies stability of form and content. Information is defined as an aggregate of data meant for communication, that is, content that is meant to be conveyed either through space or through time. A datum is the elementary unit of such content, that is, its smallest meaningful part, an indivisible unit of information. Scientists speak primarily of data while archivists preserve records. Although, from an archival point of view, data are the smallest part of a record's content. they are managed, communicated and kept in the context of other data, thereby becoming information, which, manifested in a given form and affixed to a medium, constitutes a document. This document, if set aside with other related documents for action or reference, is a record. Thus, when scientists talk about preservation of data, they are really referring to preservation of records, as there is no other way of preserving data than as records, and there is no conflict whatsoever between the preservation goal of an archivist and that of a scientist. Of course, this reconciliation of meaning for the purpose of research on authentic preservation has resulted from much analysis, several discussion papers, and a couple of scholarly papers in course of publication, but, most importantly, it is derived from and supported by the evidence provided by the case studies carried out in the scientific area.

### **2.3.3 Open inquiry**

InterPARES 2 espouses no epistemological perspective or intellectual definitions *a priori*. Instead, researchers in each working group identify the perspective(s), research design, and methods that they believe to be most appropriate to their inquiry. The reason for this openness is that InterPARES 2 (InterPARES 2 Project, 2002-2006) is conceived to work as a "layered knowledge" environment, in the sense that some of the research work builds upon knowledge developed in the course of InterPARES 1 (InterPARES 1 Project, 1999-2001), some uses knowledge developed by other teams investigating digital preservation, such as ERPANET in Europe (ERPANET, 2001-2004), some takes knowledge of similar issues developed in other areas of endeavour and bring it to bear on creation and preservation of digital materials, some reconciles knowledge about records and their attributes, elements, characteristics, behaviour and qualities existing in various disciplines and develop it for archival purposes, and some explores new issues and study entities never examined before and develop entirely new knowledge.

### **2.3.4 Multi-method design**

As stated, each research activity is carried out using the methodology and the tools that the dedicated investigating team considers the most appropriate for it. Examples of the methods used are surveys, case studies, modeling, prototyping, diplomatic and archival analysis, and text analysis.

The research is guided by detailed questions that specifically address 1) the records creation process in each of the examined areas of endeavour, and the characteristics, structure and interrelationships of the resulting materials; 2) the issues related to the development of a chain of preservation that begins with data creation and includes records evaluation and selection for permanent preservation, records description, and records reproduction as authenticating procedures; 3) the meaning of the concepts of accuracy, reliability and authenticity in the various disciplines; 4) the policies, strategies and standards in each area of activity covered by the research; 5) the descriptive schemas necessary to the identification, use and preservation of the materials produced by each activity throughout their life-cycle; and 6) the models that more appropriately represent the digital object that is investigated at any given time and the processes of its creation, maintenance, use, selection and preservation.

## **3 RESEARCH PROGRESS**

The need to concentrate the initial part of the research on gathering an understanding of the process of creation in interactive, dynamic and experiential digital environments has been especially encouraged

and supported by the participant stakeholders. The researchers have carried out case studies and general studies.

### **3.1 Case Studies and General Studies**

The case studies were identified according to the specific kind of activity that generated the material, and conducted by individual teams assembled in an interdisciplinary way for the purpose of investigating the entire life cycle of the digital objects that were examined. Each team comprised at least a scholar of the activity under investigation, a technology expert, an archival scientist, and a student research assistant. Depending on the complexity of the case study, additional experts and students might join the team. The general studies were developed to address issues relevant to each of the three types of activities producing records, but not specific to any given case.

Examples of the case studies undertaken in the scientific focus are:

- Archaeological Records in a Geographical Information System: Research in the American Southwest. This case study focuses on data sets from geographic information systems that were created by the Centre of Desert Archaeology in Tucson, Arizona. Specifically, how the systems were created, what data they contain, what their life cycle is, what happens after disposition, and their corresponding authenticity, reliability and accuracy while undergoing these processes.
- Preservation and Authentication of Electronic Engineering and Manufacturing Records. The records that are examined in this case study have been created in computer-assisted engineering, computer-assisted design and industrial automation systems. The focus of this case study is on examining the ability of complex engineering records to stand for the solid objects modeled in the records, and the ability of the manufacturing records to represent the processes required to produce such solid objects.
- The CyberCartographic Atlas of Antarctica. This project seeks to explore the new theoretical construct of CyberCartography, which sees the map as a new organizing mechanism for digital information in the information era. The dynamic, multi-dimensional, multisensory and multimedia Atlas will become an important scientific digital knowledge asset that will from its inception include archiving as an integral component of the project. The objectives of this case study are to explore the process of record creation and the function of the data within the activity in which it participates, and to determine which of a record's features will allow its authenticity to be determined.
- MOST Satellite mission. This space telescope monitors variations in the brightness of the stars. The data consist of series of nearly uninterrupted measurements of stars fields transmitted to a network of three radio ground stations. The primary repository is at the University of British Columbia. The MOST observations are available only to the MOST Science Team for a year, after which they must be made available to the public at large, distributed according to protocols that ensures accuracy both of the data and of their retrieval. The purpose of this case study is to develop the necessary protocols and to establish a procedure for releasing data to the public in a way that their authenticity can be verified.

Examples of general studies are:

- Persistent Archives Based on Data Grids. This study focuses on the San Diego Supercomputer Centre's project to develop a prototype for a persistent archives based upon data grid technology for the National Archives and Records Administration (NARA). It examines the minimal capabilities needed within grid technology for preservation of governmental records, focusing on activities related to the preservation of NARA's selected digital holdings.
- A survey of the e-science literature for file formats and encoding languages that are used for non-textual scientific data, information and records. File formats and encoding languages are also analyzed to determine data, information and/or record structure and other properties related to the concepts of accuracy, reliability and authenticity of the digital objects in question. In addition, the study will determine equivalence classes of file formats and encoding languages and identify conversion tools that can be used for migration.

Some of the documents resulting from case studies and general studies are already posted on the public section of the InterPARES website ([www.interpares.org](http://www.interpares.org)), under "InterPARES 2," "research to date", "project documents."

Several other studies are carried out by dedicated research units, which are concerned with large issues equally affecting all the disciplinary and professional areas covered by the research, such as metadata or policy. A brief overview will provide a sense of the breath of the project's investigation.

### 3.2 Terminology

A terminology team has been working toward the standardization of the vocabulary within InterPARES, so that all researchers and research assistants will be consistent in their usage of terms, thereby ensuring effectiveness of communication within the project, and consistency in dissemination activities. The work of the team has focused on the creation of four lexicographic instruments: 1) a Register of all terms and phrases used within InterPARES; 2) a Dictionary including the definitions provided for those terms by all the disciplines involved in the project; 3) a Glossary including the definition for each term chosen by InterPARES as the authoritative meaning for all the project's documents; and 4) a Thesaurus that assigns the terms to a specific facet (e.g. agent, action, object, event, property), links it to other facets through hierarchical, equivalence or associative relationships, and points to the preferred term. The images below are views of the database that can provide an idea of the way it works.

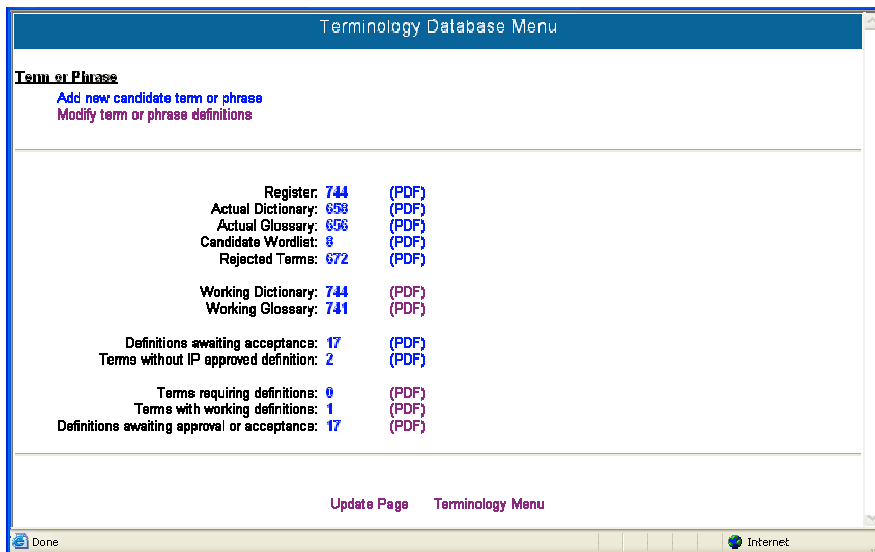


Figure 1. A view of the working database.

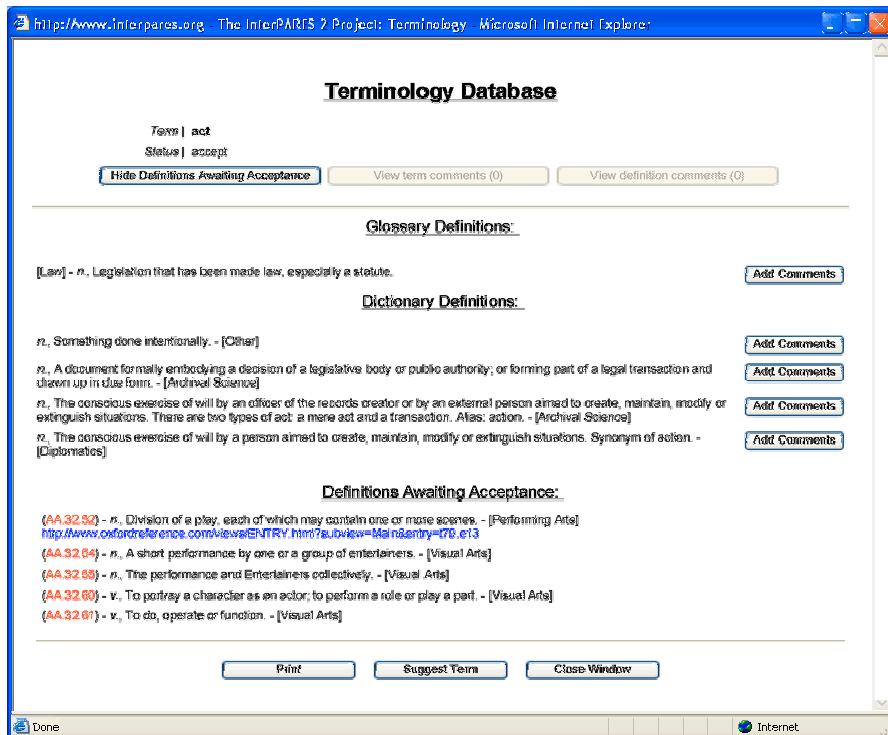


Figure 2. An example of a term from the Terminology Database.

The terminology team has worked closely with all other research units in order to ensure the greatest inclusivity while maintaining consistency and rigour, and in order to keep the delicate balance between accurately describing the nature of concepts used across the various disciplines and prescribing a common language for the research project as a whole. The team is particularly aware of the political nature of its endeavour and of the partiality and biases potentially involved in the prescription of terms and the standardization of meaning. This is a big challenge, if one considers the encompassing nature of the research, both in terms of the diversity of cultures involved, and of the multiplicity of meanings often associated with terms used in the vast range of fields incorporated within the project. In addition, InterPARES researchers have noted that key terms used by the project (e.g., accuracy, authenticity) have a multiplicity of meanings also within each of the disciplines involved in the research. As a consequence, a dedicated research unit has been assigned to gathering an understanding of the concepts linked to each of those terms within each discipline in question. This investigation has resulted in a complex and enlightening essay that will soon be published.

At present, the terminology research is focused on the evaluation of its lexicographic instruments, in order to assess their compliance with international standards, and their ability to meet the needs of external users. Future goals center on the promotion of public access to the terminology database so that researchers from various disciplines will be more aware of impediments to effective interdisciplinary communication and of the possibilities for increased consistency and understanding among disparate fields of study.

### 3.3 Modeling

The project has incorporated the use of modelling as a means to conceptually represent the activities and entities under investigation in order to assist in the analysis of case study data and to communicate findings related to the creation, maintenance and preservation of the records under examination. The major efforts of the modeling team have been channelled into the development of a Chain of Preservation model that depicts all the activities involved in the management of digital data throughout their lifecycle, from creation through selection and description to permanent preservation and access. This model integrates three pre-existing models: the model of the activities involved in managing current records produced by the University of British Columbia/US Department of Defense project in 1994-97 (Duranti, Eastwood & MacNeil 2002), and the two models of the functions of selection and preservation of digital records produced during the first phase of the InterPARES project. The latter model was based on the OAIS model, but was at a level of much higher specificity (Duranti, 2002). The modeling technique used for all models is IDEF0. The two images below respectively show how the model incorporates the entire life-cycle of records and how many modeled functions are comprised in the preservation function.

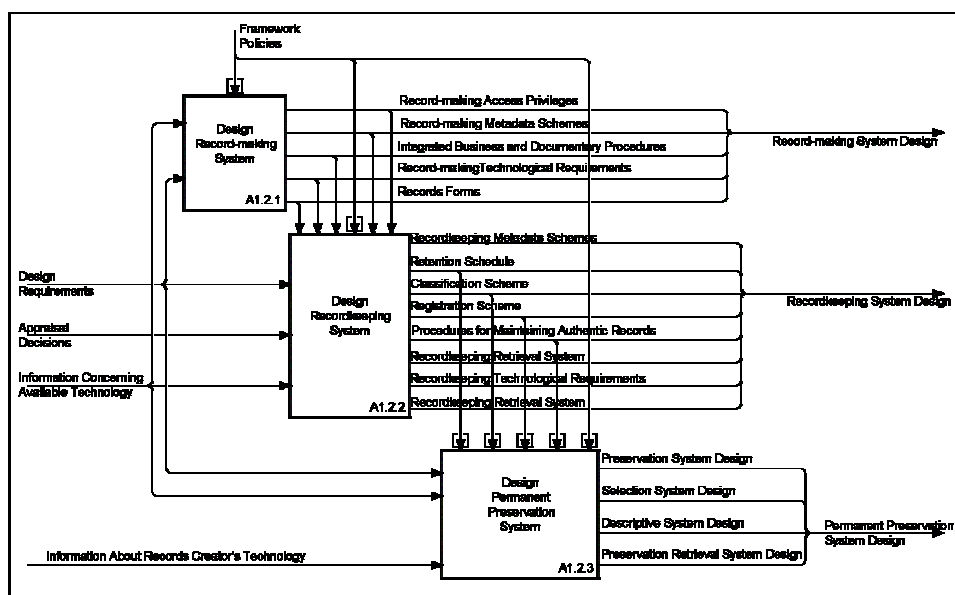


Figure 3. Activity "Design Framework" from the Manage the Chain of Preservation Model.



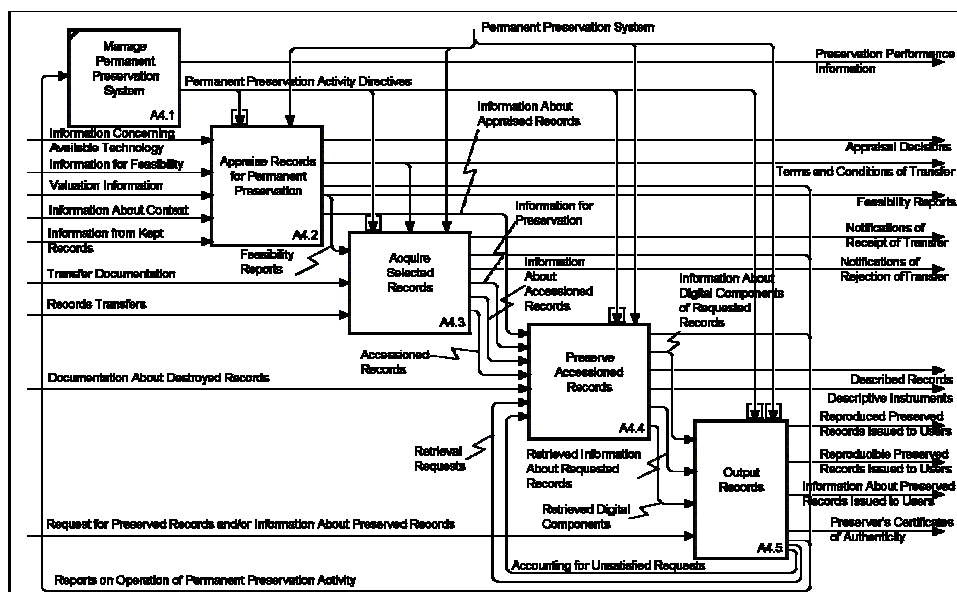


Figure 4. Activity “Preserve Selected Records” from the Manage the Chain of Preservation Model.

The Chain of Preservation model shows the relationship between the activities of the records creator and those of the records preserver and presents a clear visual representation of the ideal management process, providing the conceptual framework in which the realities of current practice may be measured and evaluated. The model assists all other research units by providing a means to identify preservation issues related to the case studies (which are walked through the model and validate it), and to assess the points of the data lifecycle that require development of policies, procedures and standards (including metadata schemas, reproduction requirements, etc.). The group is currently engaged in testing the model and has begun to develop an activity model of the business process of maintaining active data in a way that they can be preserved over the long-term. The Chain of Preservation Model is posted on the InterPARES 2 website, under “research to date” and “models”.

### 3.4 Policy

The policy research team has endeavoured to produce a conceptual framework for the development of policies that will impact each of the three foci of investigation. The major area of enquiry centres on the identification of barriers to preservation which currently exist in laws, regulations, policies and standards on copyright and intellectual rights, privacy and freedom of information, authenticity and authentication, open standards and open source, and data, information, records and archival management. Legislation and standards originating in Canada, the United States, Australia, Singapore, Hong Kong, and the European Union (as well as specific countries within it), which have a bearing on data and records creation, maintenance and preservation are being analysed within the conceptual framework in order to formulate model policies and standards. Case studies have aided in the identification of areas in which policies are absent or insufficient to ensure the preservation of authentic electronic records created in the arts, science and e-government. Preliminary findings have indicated the need to return to the articulation of the concepts of fixity and stability within each environment – an endeavour that is currently a major focus of discussion among the researchers. A series of model policies and guidelines will then be able to ensure that the entities identified as potentially fixed and stable are managed appropriately throughout their lifecycles. One objective of this team is to establish liaisons with policy and standards-issuing organizations in order to ensure their effective implementation.

### 3.5 Description

The main objective of the research team focusing on the description of digital entities has been to evaluate existing and emerging metadata schemas, descriptive standards and metadata tools to determine their ability to meet requirements relating to the creation, management, appraisal, preservation and use of accurate and authentic digital objects in the three areas. The purpose of this

endeavor has been to recommend either the development of new schemas, standards and tools, or the extension of existing and emerging ones.

The resultant activities of this research team have consisted of several specific projects. The team has begun the analysis of the completed case studies in order to identify the various uses of metadata and other descriptive standards within the diverse spheres of activity for the purposes of learning about existing practices, setting a framework for the delineation of best practices, and determining those areas in which current standards are unable to meet the requirements identified. A specific product which is intended to both facilitate this research and be a direct deliverable of the analysis is a metadata schema registry called Metadata and Archival Description Registry and Analysis System (MADRAS). MADRAS is a centralized repository of schemas, which will aid various professions and organizations to identify metadata sets, or the combinations of elements from several sets, which are appropriate to serve their data or record-keeping needs. The registry describes relevant existing and emerging recordkeeping metadata schema and archival descriptive standards in a standardized way by identifying their scope and purpose, the type of metadata/descriptive elements, the related encoding schemas, and the applicability of the schemas and standards to recordkeeping and archival functions. It analyses and evaluates schemas and standards in relation to recordkeeping and archiving requirements as set down in standards. The metadata schema for the registry itself includes 120 fields organized hierarchically in an element structure that is relatively flat, going 3 levels deep. The first level of the hierarchy comprises eleven elements: Registration, Identification, Accessibility, Rights, Provenance, Description, Analysis, Documentation, Relationships, Administration, and a general Note element. The image below shows one view of MADRAS.

**InterPARES2 Metadata and Archival Description Registry and Analysis System (MADRAS) BETA VERSION**

[Home](#)

**Browse Schemas**

- Summary View
- Identification >>
- Description
- Documentation
- Implementation
- Rights
- Provenance
- Relationships
- Software
- Note
- Analysis Document

[Guidelines](#)

[About Us](#)

**BUG FIXES & UPDATES**

1. user\_provenance: when submitting a new agent, it is overwritten with State Records Authority of New South Wales.  
fix: pending

**AA\_ISO19115 Geospatial Metadata Standard Final Draft**

Data elements to identify and distinguish metadata schemas.

IDENTIFICATION		
Unique Global ID	Domain	URL
	ID	www.isotc211.org/ smXML; www.isotc211.org/sxXML
Official Name		AA_ISO19115 Geospatial Metadata Standard
Alternative Name	Acronym	-----
	Previously Known As	ISO 15046-15
	Other	
Version		Final Draft
Status		draft
Country of Origin		The specifications of the ISO 19115 are based on
Publication	Publisher	Prepared by ISO/TC 211, Published by American National Standards Institute,
	Place	Published in Switzerland
	Date	2003
Duration	Start Date	??
	End Date	??
Note	1. Unique Global ID is not specified in ISO 19115. I am not sure if I can use the xml name space identifier from the implementation standard.	

Figure 5. View from the MADRAS database: the identification page of a registry entry.

At this time, the design phase of the database prototype is completed, with implementation to occur by the end of 2005, when the production version will be posted on the InterPARES public site. The environment of the production version is a MySQL database, PHP script, and Apache web server running on Windows. The website comprises a public area where completed schemas are published, and a private area where researchers create and manage schema records. The system allows for online distributed collaboration, authentication of users and basic workflow management. Approximately twenty-five of the most relevant metadata schemas have been analyzed in reference to ISO recordkeeping standards, InterPARES requirements for the presumption of the authenticity of records, and the Australian Recordkeeping Metadata Schema, while many more schemas have been identified for registration and analysis (Gilliland-Swetland, 2003; Turner, 2004).

Initiated by the description team, but built upon by all InterPARES research teams, a literary warrant database has been developed in order to facilitate the identification of authoritative sources relating to the ways in which metadata and archival description support data/record creation and preservation activities. This database has been implemented and is currently being populated with references from these sources. The description team is also working closely with the modeling team to identify the requirements for the creation of metadata, the specific points within the Model of the Chain of Preservation at which they must be created, and the responsibilities for such creation.

## **4 CONCLUSION**

The InterPARES 2 Project has already produced a large quantity of the material on the basis of which it will develop the project's deliverables, that is, among other things, guidelines for data and records creators outlining methods for the reliable production and maintenance of data that can be authentically preserved; prototypes of appraisal and preservation systems, and guidelines for data and records preservers; frameworks for developing policies, strategies and standards, and for the development of descriptive standards for the digital objects under examination; registries of metadata schemas; and literature and terminology databases. However, it has been noted by several observers that the most desirable outcome of InterPARES has already been achieved: the harmonious collaboration of scholars and professionals from such a large variety of disciplines, backgrounds and cultures towards the long-term preservation of their digital memory is the invaluable product of InterPARES that all team members and most stockholders value the most.

## **5 REFERENCES**

- Brown, R. (2000). Virtual unreality and dynamic form: an exploration of space, time and energy. *Leonardo* 33 (1), 21-25.
- Boudrez, F. & Dekeyser, H. (2003). DAVID: *Digital archiveren in de Praktijk. Handboek*. Antwerpen-Leuven: Interdisciplinary Centre for Law and Information Technology and Stadsarchief Antwerpen.
- Council on Library and Information Resources. (2000). *Authenticity in a Digital Environment*. Washington, D.C.: CLIR.
- Chen, S. (2003). Preserving Digital Records and the Life Cycle of Information. In Zelkowitz, M. (Ed.). *Advances in Computers, Vol. 57, Information Repositories*. Maryland: Academic Press.
- Dale, R. (1998). Integrating natural language generation and hypertext to produce dynamic documents. *Interacting With Computers* 1, 109-155.
- Dollar, C. (1999). *Authentic Electronic Records: Strategies for Long-Term Access*. Chicago: Cohasset Associates.
- Downes, E. J. & McMillan, S.J. (2000). Defining Interactivity. *New Media and Society* 2 (2), 157-179.
- Duranti, L. (Ed.) (2002). *The Long-term Preservation of Authentic Electronic Records: Findings of the InterPARES Project*. Available from: <http://www.interpares.org/book/index.cfm>. Also: (2005) San Miniato, Italy: Archilab.
- Duranti, L, Eastwood, T. & MacNeil, H. (2002). *Preservation of the Integrity of Electronic Records*. Dordrecht: Kluwer Academic Publishers Group.
- Eastwood, T. (2004a). Appraising Digital Records for Long-term Preservation. *Data Science Journal* 3 (30 December), 202-208.
- Eastwood, T. (2004b). Appraising Digital Materials for Preservation as Cultural Heritage. *Proceedings of Symposium 2003 – Preservation of Electronic Records: New Knowledge and Decision-making, 2003*. Ottawa: Canadian Institute for Conservation.

- ERPANET (Electronic Resource Preservation and Access Network) (2001-2004). Homepage of ERPANET. Available from: <http://www.erpanet.org>.
- European Commission. (2000). *European citizens and electronic information: the memory of the Information Society: Proceedings of the DLM Forum on electronic records. Brussels, 18-19 October 1999*. Luxembourg: European Commission.
- Gilliland-Swetland, A. (2000). *Enduring Paradigms, New Opportunities: The Value of the Archival Perspective in the Digital Environment*. Washington, D.C.: Council on Library and Information Resources.
- Gilliland-Swetland, A. (2003). Metadata – Where Are We Going? In Gorman, G.E (Ed.) *International Yearbook of Library and Information Management 2003: Metadata Applications and Management*. London: Facet Publishing.
- Guercio, M. (2002). *Archivistica informatica: i documenti in ambiente digitale*. Rome: Carocci.
- Hodge, G. (2000). Digital archiving bringing stakeholders and issues together. A report on the ICSTI/ICSU Press workshop on Digital Archiving. ICSTI Forum 33 (March). Retrieved 3 July, 2005 from: <http://www.icsti.org/icsti/forum/33/#Hodge>.
- Hofman, H (Ed.). (2003). *Blijvend in Business. Naar een geordende en toegankelijke staat van informatie*. The Hague: Nationaal Archief.
- Huhtamo, E. (1999). From Cybernation to Interaction: A Contribution to an Archaeology of Interactivity. In Lunenfeld, P. (Ed.) *The Digital Dialectic: new essays on new media*. Cambridge, Mass.: MIT Press.
- InterPARES (International research on Permanent Authentic Records in Electronic Systems) 1 Project (1999-2001). Homepage of the InterPARES 1 Project. Available from: [http://www.interpares.org/ip1/ip1\\_index.cfm](http://www.interpares.org/ip1/ip1_index.cfm).
- InterPARES (International research on Permanent Authentic Records in Electronic Systems) 2 Project (2002-2006). Homepage of the InterPARES 2 Project. Available from: [http://www.interpares.org/ip2/ip2\\_index.cfm](http://www.interpares.org/ip2/ip2_index.cfm).
- Levy, D. (1999). The Universe Is Expanding: Reflections on the Social (and Cosmic) Significance of Documents in a Digital Age. *Bulletin of the American Society for Information Science* 25 (4), 17-20.
- Lynch, C. (2000). Experiential Documents and the Technologies of Remembrance. In Scammell, A. (Ed.) *I in the Sky: Visions of the Information Future*. London, England: Library Association Publishing.
- MacNeil, H. (2000). *Trusting Records. Legal, Historical and Diplomatic Perspectives*. Dordrecht: Kluwer Academic Publishers.
- MacNeil, H. (2002). Providing Grounds for Trust II, *Archivaria* 54 (2), 24-58.
- Moore, R. (2004). Preservation Environments, *NASA/IEEE MSST2004 Twelfth NASA Goddard/Twenty-First IEEE Conference on Mass Storage Systems and Technologies*. Washington, D.C.
- National Research Council (US). (1997). Committee on Issues in the Transborder Flow of Scientific Data. *Bits of Power: Issues in Global Access to Scientific Data*. Washington, D.C.: National Academy Press.
- National Research Council (US). (2000). *The Digital Dilemma: Intellectual Property in the Information Infrastructure*. Washington, D.C.: National Academy Press.
- National Research Council (US). (2001a). *Issues for Science and Engineering Researchers in the Digital Age: Office of Special Projects*. Washington, D.C.: National Academy Press.

National Research Council (US). (2001b). Committee on the Preservation of Geoscience Data and Collections. Committee on Earth Resources. *Geoscience Collections and Data: National Resources in Peril*. Washington, D.C.: National Academy Press.

Rokeby, D. (1995). Transforming Mirrors: Subjectivity and Control in Interactive Media. In Penny, S. (Ed.) *Critical Issues in Electronic Media*. New York: State University of New York Press.

Ross, S. (2000). *Changing Trains at Wigan: Digital Preservation and the Future Scholarship*. London: NPO Preservation Guidance Occasional Papers.

Stromer-Galley, J. (2004). Interactivity-as-Product and Interactivity-as-Process *The Information Society* 20, 391-394.

Turner, J. (2004). *Homepage of MetaMap*. Available from:  
<http://mapageweb.umontreal.ca/turner/meta/english/metamap.html>.