→ **PV 2013**

Ensuring the Long-Term Preservation and Value Adding to Scientific and Technical Data

© ESA 2013

4–6 November 2013 | ESA–ESRIN | Frascati (Rome), Italy

ABSTRACT BOOK

Preservation Begins At Creation:
A New Perspective on Digital Preservation Education

Adam Jansen [(1)], Luciana Duranti [(2)]

[(1)] **University of British Columbia**
470-1961 East Mall Vancouver, British Columbia Canada   V6T 1Z1
E-Mail: adamj@mail.ubc.ca

[(2)] **University of British Columbia**
470-1961 East Mall Vancouver, British Columbia Canada   V6T 1Z1
E-Mail: Luciana@mail.ubc.ca

ABSTRACT

This paper presents a new approach to digital preservation education; one that focuses on preservation through the *creation* of reliable, accurate, and usable digital records that are maintained as authentic over the long term. InterPARES, in collaboration with the International Council on Archives, has developed a series of educational modules titled *Digital Records Pathways: Topics in Digital Preservation* in order to assist organizations in bridging the digital preservation competency gap. Consisting of eight independent modules, *Digital Records Pathways* addresses the theoretical and practical knowledge needed to establish the framework, governance structure and systems required to manage and preserve digital records throughout the records' lifecycle. Based on findings of the InterPARES project (e.g. the *Chain of Preservation* model, Creator Guidelines and Preserver Guidelines) and the concepts and methods of records management, archival science, diplomatics, and digital forensics, these educational modules are designed to provide guidance to organizations when issuing policies, guidelines and regulations related to the creation, management and preservation of digital records.

Keywords: Archival Education, Digital Preservation, InterPARES, Authenticity, Diplomatics

INTRODUCTION

Digital records must be carefully managed throughout their entire life "to ensure that they are accessible and readable over time with their form, content and relationships intact to the extent necessary for their continuing trustworthiness as records" [1]. From an archival science perspective, the term trustworthy refers to the ability of a record to accurately reflect the activities in which it participated and evidence that those records have remained untainted by the "distorting influence of time, bias, interpretation or unwarranted opinion on the part of the record-maker" [14] or subsequent custodians. As such, preservation of trustworthy digital records begins at creation of the records and "must be thoroughly documented as a primary means for protecting and assessing authenticity over the long term. Since preservation begins at creation, responsibility for this through documentation rests with both the creator and the preserver" [15].

Records that are created in the usual and ordinary course of business, until proven otherwise, are extended a presumption of trustworthiness and they are expected to be "intact and what they purport to be …by controlled procedures of transmission and preservation, [and] can be presumed to be truthful as to their content" [2]. A trustworthy record possesses the qualities of authenticity, reliability and accuracy, while preserving the content, context, and structure of the record in an unbroken chain of legitimate custody; for maintaining a record "…without capturing such critical evidence is equivalent to destroying the record" outright [1]. Therefore, preserving trustworthy records requires knowledge of and control over the procedures that a record was involved in throughout its lifecycle.

For fifteen years, the International Research on Permanent Authentic Records in Electronic Systems (InterPARES) Project has conducted extensive research into developing and articulating the concepts,

principles, criteria and methods required to preserve trustworthy digital records. Throughout the course of the research project, one finding stood out: *Preservation Begins at Creation* [5]. Given the rapid pace of technological advancement and the corresponding obsolescence factor, waiting until records are transferred into the custody of the archives to begin the process of preservation places a tremendous, and often insurmountable, burden on the preserver. In order to ensure records' reliability, accuracy, authenticity and accessibility over time, organizations need to address the requirements for preservation beginning with a record's creation; yet within most organizations the knowledge needed to address preservation of digital records is generally lacking. Closing this knowledge gap is complicated by the lack of education resources available to organizations that are focused on the development of digital preservation strategies. To address this lack of educational resources, the International Council on Archives – Section on Archival Education and Training (ICA-SAE) collaborated with InterPARES to produce a series of educational modules based on the findings of the InterPARES Project.

InterPARES Project

Beginning in 1999, the InterPARES Project (IP1) built upon the hypotheses developed by the UBC-MAS (University of British Columbia - Masters of Archival Studies) Project that

> express[ed] the necessary and sufficient components of a complete, reliable, and authentic electronic record. These hypotheses, which have been articulated in templates, constitute the conceptual basis for establishing, firstly, whether a given electronic system contains records and, secondly, whether these records can be considered reliable and authentic [3].

The goal of IP1 was to identify those elements that are common to *inactive* digital records selected for permanent preservation that existed in administrative or legal systems, and then once identified, to develop the methodology to verify the inactive records' authenticity over time [10]. Employing a theoretical-deductive approach to identify the key elements, as well as an empirical-inductive approach to collect and analyze data from case studies, the project added to the findings of the UBC-MAS Project and expanded upon the core elements considered necessary and sufficient for preserving the authenticity of digital records.

InterPARES 2 (IP2), 2002-2007, continued the research into the concept of authenticity while focusing on the diversity of digital records contained in interactive, experiential, and dynamic environments. Incorporating the findings from IP1 along with extensive case studies, the project developed a series of reports for each of the different domain environments and focuses. To educate creators and custodians of digital records, the project developed the Creator Guidelines [12] and Preserver Guidelines [13] to assist individuals and organizations in making informed decisions with regards to the creation, maintenance and preservation of trustworthy digital records.   The project also created the Chain of Preservation (CoP) Model [4] that illustrates the chain of activities a record participates in from its creation through preservation. Analysis of this model, along with the narrative of the activity flows, reinforces the concept that the preservation of trustworthy digital records must begin at the creation of that record.

The third phase of the InterPARES Project (IP3), completed in 2012, built upon the knowledge acquired in the first two phases by translating "the theory and methods of digital preservation drawn from research to date into concrete action plans for existing bodies of records that are to be kept over the long term by archives and archival/records units within organizations endowed with limited resources" [11] in order to "enable small and medium sized public and private archival organizations and programs… to preserve over the long term authentic records" [10]. The third phase of the project put into practice the research of the earlier phases through the development of model policies and procedures based on findings from test beds that could apply the research findings to active records systems. The findings from the case studies found that the implementation of the concepts, principles, and methods were highly dependent on technological context.

International Council on Archives – Section for Archival Education and Training

The International Council on Archives (ICA) is "dedicated to the effective management of records and the preservation, care and use of the world's archival heritage through its representation of records and archive professionals across the globe" [6]. A neutral, non-governmental body with over 1400 participating members from 199 countries and territories, the ICA has been active over the past 60 years in supporting archivists and archival institutions in advocating for the proper management and protection of records of enduring value, production of archival standards and development of best practices [7]. The mission of ICA's Section for Archival Education and Training (ICA-SAE) "aims to establish closer contacts among teachers of archival science and to work towards the development of methodology and preparation of tools to assist them in their teaching tasks and in their contributions to the development of the profession" [8]. In support of that mission, ICA-SAE collaborated with the InterPARES to produce a series of education modules titled *Digital Records Pathways: Topics in Digital Preservation* based on the research findings of InterPARES – such as the Chain of Preservation Model, Terminology Database, Creator Guidelines and Preserver Guidelines.

Chain of Preservation Model

The Chain of Preservation (CoP) model, developed during the course of IP2, documents the activities, inputs, outputs, constraints and mechanisms that are involved in the process of creating, managing and preserving trustworthy records. The CoP model itself is built upon three earlier models: the creator-centric preservation model developed by the UBC-MAS project, and the two preserver-centric models of the selection function and the preserver function developed during IP1. The CoP model, in the process of integrating and expanding the three previous models, took a record-centric approach by focusing on the needs of the record, and supported the model with a glossary of terms to provide clarity as needed. The CoP model demonstrates that preservation of digital records requires a comprehensive understanding of all phases of a record's lifecycle from its point of creation through designation for permanent preservation. This approach is reflected in the name chosen for the model – Chain of Preservation. As in a real chain, all activities that a record participates in are permanently intertwined, or linked together. Any weakness, omission or compromise in any one of the links deteriorates that chain's ability to perform its function. As such, the preservation of trustworthy digital records must take into account all the activities that a record participates in starting with creation; a lack of understanding in any given phase compromises a custodian's ability to ensure that the records will be accessible, readable, usable and trustworthy over time.

To produce a highly detailed model of the activities involved in the creation, management and preservation of digital records, InterPARES chose to create the model using the Integrated Definition Language (IDEF) [4]. The IDEF modeling language is "useful for the purposes of analyzing and graphically representing the diplomatic and archival concepts, and making their meaning comprehensible and relevant to system designers," [3] by documenting the inputs, outputs, constraints, mechanisms for each activity in a hierarchical relationship. As understood by the model:

- Inputs represent those pieces of information or objects that are involved in the process but originate from outside the activity being defined. These inputs can come from other activities or from external sources not involved in the process prior to that activity;
- Outputs are the result of the activity, typically in the form of the records themselves or additional information about the records that was discovered and/or created as a result of the activity. These outputs can stand on their own, and as such much be preserved along with the records for the entirety of the record's lifecycle, or are inputs into activities further down the process chain;
- Mechanisms are resources that help to create, manage, control and preserve the digital records. These resources can take the form of technology, equipment and/or personnel that are necessary to complete a given activity. The omission of any mechanism can become a constraint on the activity, up to and including the inability of that activity to occur; and,

Constraints on the records vary by the context of the creation and maintenance of the record, as well as the particular phase of the lifecycle that the record is in at the time. These constraints differ from organization to organization and are influenced by such factors as the juridical system within which the organization operates including applicable laws and regulations, the technology used, and the national and international standards for operation.
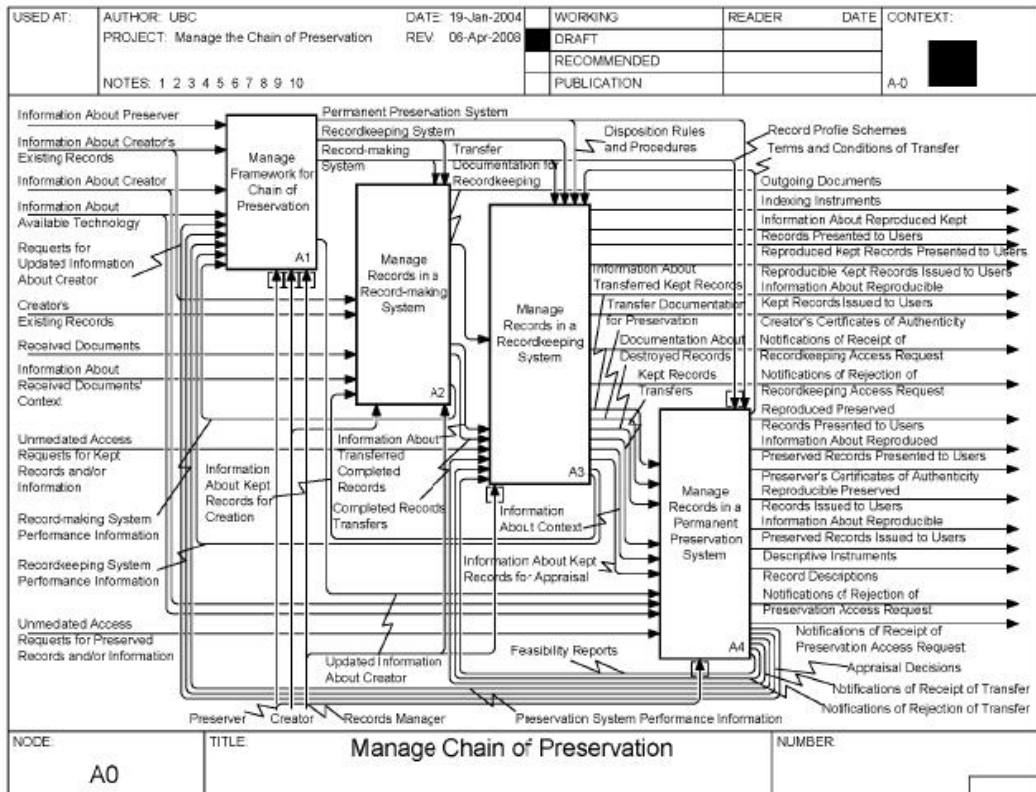

Figure 1: Example of an activity from the Chain of Preservation (CoP) Model [4]

The CoP model focuses on four primary records activities: managing the framework for the chain of preservation, managing records creation, managing records within a recordkeeping system, and preserving selected records. Each of these primary activities is then decomposed into the sub-activities that are necessary to accomplish the task, with each sub-activity being further decomposed as necessary along with each activity's respective inputs, outputs, mechanisms and constraints. By decomposing the activities and documenting the parameters involved in that activity, the CoP model explicitly illustrates the interrelation between the records activities. By following the series of diagrams for any given activity and its accompanying narrative, an institution can assess their readiness to preserve trustworthy records by comparing the inputs, outputs, mechanisms and constraints available within their institution against the model developed by InterPARES. This assessment provides an effective framework for guiding preservation decisions and implementing the appropriate policies and procedures needed to support the preservation of trustworthy digital records.

The Educational Modules

In 2010, ICA-SAE and InterPARES began the educational initiative to provide digital preservation education and training to archivists, records managers and archival educators based on the InterPARES findings that became the *Digital Records Pathways: Topics in Digital Preservation* educational modules. The architecture of the modules was designed to begin with foundational knowledge concepts covering the framework for digital preservation and developing policies regarding preservation; progress into more general knowledge about records through understanding of an organization's culture, the role metadata plays in preservation, and the methods for moving from an uncontrolled records environment into a

structured one; and conclude with specific technical knowledge about managing e-mail, records on websites and Cloud Computing.

The modules were designed to progress from a broad subject area into specific technical topics in order to provide a strong overview of digital preservation, with each independent module also standing on its own for individuals already possessing a strong foundational knowledge and only interested in specific modules. The modules rely upon a series of external resources to supplement each module's content and to address any knowledge gaps on the part of the learner that might exist. While the modules were developed by the InterPARES Canadian team and contain many examples and case studies drawn from Canadian institutions, the modules themselves were designed to be expanded and adapted to reflect the context of the intended audience and to be updated regularly as technology changes.
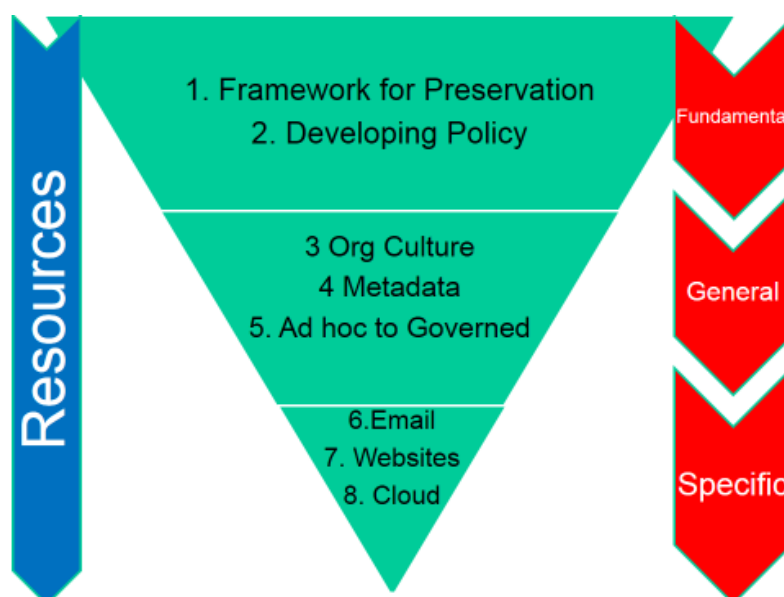

Figure 2: Architecture Overview of Education Modules

The eight modules in the series [9] are:

- *Module 1: Framework for Preservation* which provides an overview of the modules in the series; resources for institutional self-assessment and readiness; and, highlights of two complimentary models concerning digital preservation – the Open Archival Information Systems (OAIS) model and the Chain of Preservation (CoP) model;
- *Module 2: Developing Policies and Procedures* which details the purpose and benefit of digital preservation policies; and provides the knowledge and tools necessary to write policies and procedures that will support the preservation of trustworthy digital records;
- *Module 3: Organizational Culture* which discusses organizational cultural as either an enabling or constraining factor in proper recordkeeping; outlines the different types of organizational culture likely to be encountered in organizations; provides tools, such as checklists and a list of indicators, for stakeholders to assess the types of culture within their unit and their organization; and strategies to promote records management based on the type of culture within their organization;
- *Module 4: Metadata* which provides an overview of the importance of the role of metadata in proper records management and preservation; categorizes metadata into descriptive, administrative, and technical categories; presents different methods of representing or encoding metadata, as well as a primer on common metadata standards; and introduces the InterPARES Application Profile for authenticity metadata;

- *Module 5: Appraisal Strategy for Network Drives* which discusses the appraisal of digital records as four distinct activities – compiling information about the records, assessing the value of the records, determining the feasibility of preservation the records, and making the decision to acquire the records; frames the assessment of the authenticity of records as an integral part of the appraisal process through the use of the Benchmark Requirements and appraisal guidelines for legacy records; and presents an action plan for moving records from unstructured file stores to controlled recordkeeping environments;
- *Module 6: E-mail Management* which focuses on gaining control of e-mail by implementing new policies and procedures; presents the E-mail Management and Preservation Model (EMPM); discusses factors that influence e-mail management and preservation; and suggests methods for managing e-mail that include applying retention to e-mail, preserving e-mail, and designing supporting policies and procedures to support proper records management;
- *Module 7: Preservation of Web Records* which discusses the management and preservation of records stored in the web; provides methods for identifying records amongst all the non-records on websites; discusses analyzing management and preservation needs for web-based records; suggests a workflow management process for managing creation and movement of records to and from the web; and provides a policy framework for an organizational approach to web record management; and,
- *Module 8: Cloud Computing* which presents a primer on Cloud Computing; presents records management issues to consider before moving records to the Cloud; defines what the Cloud is; provides tools for analyzing risk when migrating records to the Cloud; and provides a foundation for developing an organizational Cloud Computing strategy that includes user education and business requirements, rules and compliance.

CONCLUSION

The preservation of trustworthy digital records is a complex task that must account for the activities that created, managed and maintained those records throughout their lifecycle.   Extensive research by the InterPARES Project into preserving trustworthy digital records has clearly demonstrated that such an endeavor begins when a record-making system is designed. The custodian of the records must be able to show that the processes that were used to create, manage and maintain the records were trustworthy and that the record can be accurately reproduced over time. This requires both a consistent and reliable set of policies and procedures that begin with the creation of the record, as well as documentation of the chain of custody of the record throughout its lifecycle. The technical capability to preserve trustworthy digital records currently exists, but the knowledge, skills and abilities of the custodians of those records is lagging behind. It is our hope that the ICA-SAE and InterPARES education modules on digital preservation will help filling this knowledge gap. The educational modules should "provide archivists and records professionals with the necessary theoretical knowledge and procedural and strategic skills to develop, implement and monitor a digital recordkeeping and/or preservation system" [15]. By working through the series of education modules, custodians of digital records will learn an effective, tested framework to implement the procedural and technological changes necessary to preserve trustworthy digital records of permanent fiscal, historical, legal, cultural and intellectual value.

REFERENCES

[1] - Bearman., D. (1993) The implications of Armstrong v. Executive Office of the President for the archival management of electronic records. *American Archivist* 56, 150-160.

[2] - Duranti, L. (1995). Reliability and authenticity: The concepts and their interpretations. *Archivaria* 39, 5-10.

[3] - Duranti, L., & MacNeil, H. (1996). The protection of the integrity of electronic records: An overview of the UBC-MAS research project. *Archivaria* 42, 46-67.

[4] - Eastwood, T., Ballaux, B., Mills, R., & Preston, R. (2008). Appendix 14: Chain of preservation model diagrams and definitions. In L. Duranti, & R. Preston (Eds.), *International research on permanent authentic records in electronic systems (InterPARES) 2: Experiential, interactive and dynamic records* (pp. 454-549). Padova, Italy: Associazione Nazionale Archivistica Italiana.

[5] - Hackett, Y. (2008). Part four: Methods of appraisal and preservation --domain 3 task force report. In L. Duranti, & R. Preston (Eds.), *International research on permanent authentic records in electronic systems (InterPARES) 2: Experiential, interactive and dynamic records* (pp. 161-192). Padova, Italy: Associazione Nazionale Archivistica Italiana.

[6] - International Council on Archives. *ICArchives : Presentation of the ICA : Discover ICA*. Retrieved 14 Oct 2013, from http://www.ica.org/104/presentation-of-the-ica/discover-ica.html

[7] - International Council on Archives. *ICArchives : About ICA : An introduction to our organization*. Retrieved 14 Oct 2013, from http://www.ica.org/102/about-ica/an-introduction-to-our-organization.html

[8] - International Council on Archives – Section for Archival Education and Trainin. *ICA-SAE.* Retrieved 14 Oct 2013, from http://www.ica-sae.org/mission.html

[9] - InterPares and International Council on Archives. *Digital records pathways: Topics in digital preservation -- Module 1: Introduction – A framework for digital preservation.* Retrieved 08 Oct 2013, from http://www.ciscra.org/docs/Education_Modules_Digital_Records_Pathways.zip

[10] - InterPARES. *InterPARES project summary.* Retrieved, 24 Sept 2012, from http://www.interpares.org/ip1/ip1_index.cfm

[11] - InterPARES. *InterPARES 3 project overview.* Retrieved 20 Sept 2012 from http://www.interpares.org/ip3/ip3_overview.cfm

[12] - InterPARES. *Creator Guidelines – Making and Maintaining Digital Materials: Guidelines for Individuals.* Retrieved 6 Oct 2013, from http://www.interpares.org/ip2/creator_guidelines/creator_guidelines.html

[13] – InterPARES. *Preserver Guidelines – Preserving Digital Records: Guidelines for Organizations.* Retrieved 6 Oct 2013, from http://www.interpares.org/ip2/preserver_guidelines/preserver_guidelines.html

[14] – MacNeil, H. (2001). Trusting Records in a Post-Modern World. *Archivaria* 51, 36-47.

[15] - Rennie, S. (2009) Bridging time: InterPARES and e-dhiscovery. Retrieved 12 Oct 2013, from http://www.law.pitt.edu/DESI3_Workshop/Papers/DESI_III.StuartRennie.pdf

Acknowledgments