Do You (Still) Have the Real Thing? Using the InterPARES 2 *Framework of Principles* to Address Authenticity in Preservation Process Assessments

Randy Preston; InterPARES Project, The University of British Columbia; Vancouver, BC, Canada

Abstract

Preservation process assessments assist records creators and preservers in understanding the complex regime of internal and external factors that affect the long-term care of the records under their care, while at the same time highlighting problem areas. Because of the distributed nature of digital preservation, to be effective, these assessments need to take into account issues related to the entire chain of custody, from creation through preservation, so that the *authenticity* of the records is maintained throughout. Doing so will require that creators and preservers develop new strategies for instituting and sustaining more active and integrated records management collaborations supported by a comprehensive and *harmonized* intellectual framework of policies, procedures, practices, and standards.

The purpose of this paper is twofold: (1) to introduce to professional practitioners involved in digital preservation activities, especially practitioners outside the archival community, the concept of authenticity as it has been developed during the past eight years of intensive research by the InterPARES (International Research on Permanent Authentic Records in Electronic Systems) Project, and (2) to examine how the findings of this research, particularly the InterPARES 2 *Framework of Principles*, and the intellectual framework of model policies, principles and standards it supports, can help encourage and support more comprehensive and integrated preservation process assessments aimed at improving the ability of creators and preservers to establish and maintain the authenticity of the digital records, and other digital content objects, under their care.

Introduction

Preservation process assessments assist records creators and preservers in understanding the complex regime of internal and external factors that affect the long-term care of the records under their care, while at the same time highlighting problem areas. By assessing, in as holistic a manner as possible, the nature of the materials requiring preservation against the organization's relevant policies, procedures, practices, stakeholder relationships, and technological and physical infrastructure and resources, long-term preservation needs are determined and prioritized, and the human, capital and intellectual resources required for implementation are identified. The primary aim of the assessment is to provide a comprehensive review of current circumstances and projected future needs and risks so that the organization can develop an informed and effective preservation plan. Among other things, the findings of such an assessment can help establish and/or support effective records creation, maintenance and preservation policies and procedures, increased organization-wide awareness of records management issues, and reallocation of existing resources to better harmonize the full spectrum of records management activities in support, ultimately, of effective long-term preservation of authentic records. If the assessment methodology is sound, reliable and sensitive to the unique needs and constraints of the organization, and the organization's efforts at implementation are continuous and supported by all levels of administration, the result is a comprehensive preservation program that provides the framework for effectively and efficiently addressing the organization's long-term preservation requirements.

For at least the past three decades, researchers and practitioners such as Cunha, founder of the Northeast Document Conservation Center (NEDCC), have emphasized the importance of ensuring that preservation institutions, such as libraries, museums and archives, consciously, systematically and routinely (re)assess their preservation needs. [1] It has only been within the past decade or so, however, that any serious effort has been made to include "intangible" digital media in these preservation assessments. [2] Despite an initial naïveté about characterizing the problems inherent in digital preservation as primarily technological (e.g., media fragility, technological obsolescence), it is now clear that technological concerns are but one part of a far more complex and nuanced preservation puzzle, and that any viable, long-term solution will require (at least) as much emphasis on non-technological variables, such as organizational process, policy and socio-cultural issues, as on technological issues. [3]

Due to the relative ease with which most digital content objects can be accidentally or surreptitiously accessed, copied, altered and instantaneously transmitted to other computers both internal and external to the creator's or preserver's institutional domain, establishing and maintaining the authenticity of digital content objects has emerged as one of the most fundamental non-technological concerns affecting digital preservation efforts. Indeed, the crux of the authenticity issue was perhaps most poignantly and succinctly characterized by former Newbery Library president, Charles T. Cullen, when, during a January 2000 workshop organized by the Council on Library and Information Resources (CLIR), at which a panel of information experts was attempting to clarify the meaning of authenticity in relation to digital content objects, he mused, "why preserve what is not authentic?" [4] Related to this question was concern over what many at the workshop characterized as a general under-appreciation of the complexity of the tasks required to establish and maintain the authenticity of a digital content object. Together with the absence of any readily available means for testing the authenticity of a digital content object, this fueled (and continues to fuel) a more fundamental concern that the mere act of preserving an object in digital format will be seen by unwary users

of that object as implying "an endorsement of authenticity, even if nothing else is done to it," despite the fact that, as a general rule, "digital objects bear less evidence of authorship, provenance, originality, and other commonly accepted attributes than do analog objects." [5]

With these issues in mind, the purpose of this paper is twofold: (1) to introduce to professional practitioners involved in digital preservation activities, especially practitioners outside the archival community, the concept of authenticity as it has been developed during the past eight years of intensive research by the InterPARES (International Research on Permanent Authentic Records in Electronic Systems) Project, and (2) to examine how the findings of this research, particularly the InterPARES 2 *Framework of Principles*, and the intellectual framework of model policies, principles and standards it supports, can help encourage and support more comprehensive and integrated preservation process assessments aimed at improving the ability of creators and preservers to establish and maintain the authenticity of the digital records, and other digital content objects, under their care. The paper begins with a brief introduction to the InterPARES Project, highlighting the key findings and products most relevant to the discussion at hand. This is followed by an examination of the concept of authenticity, especially as articulated by the InterPARES research. A brief summary of the work of the Policy Cross-domain then introduces the Project's *Framework of Principles* document, which, as is discussed in the remainder of the paper, can serve as the overarching framework for guiding and managing preservation assessment processes through the development of model policies, strategies and standards for the long-term preservation of authentic digital records.

The InterPARES Project

The InterPARES Project, which officially concluded its research activities in December 2006, was a collaborative international research project involving more than 100 researchers, spanning 21 countries and five continents, from such diverse fields as archival science, diplomatics and records management; music theory, composition and performance; film theory, production and description; dance and theatre theory; a variety of hard and social sciences; jurisprudence; industry; government and public administration; and computer science and engineering. The Project aimed at developing the theoretical and methodological knowledge essential to the long-term preservation of authentic records created and/or maintained in digital form. As is stated on the InterPARES Web site, [6] the project was developed in two phases:

> InterPARES 1 was initiated in 1999 and concluded in 2001. It focused on the preservation of the authenticity of records created and/or maintained in databases and document management systems in the course of administrative activities. In addition, a component of the project was dedicated to the exploration of the issues related to the long-term preservation of digital sound, the findings of which led to InterPARES 2.
>
> InterPARES 2 was initiated in 2002 and concluded in 2006. In addition to dealing with issues of authenticity, it delved into the issues of reliability and accuracy from the perspective of the entire lifecycle of records, from creation to permanent preservation. It focused on records produced in complex digital environments in the course of artistic, scientific and e-government activities.

As is outlined in this paper, the knowledge generated by both phases can be used to provide the basis from which to formulate model policies, strategies and standards capable of ensuring the longevity of digital records and the ability of users to trust the authenticity of those records. To this end, InterPARES has developed a number analytical instruments and tools aimed at helping both individuals and organizations manage the creation, maintenance and long-term preservation of authentic digital records (while the focus of these tools is on digital records, they are in fact scalable to all digital content objects).

One of the key tools developed by InterPARES 1 is the **Template for Analysis**, which essentially is a decomposition of a digital record into its four necessary constituent parts: documentary form (i.e., intrinsic and extrinsic elements), annotations, contexts (i.e., the framework of action in which the record participates, including its administrative, provenancial, procedural, documentary, and technological contexts), and medium. [7] The Template defines each element, explains its purpose, and indicates whether, and to what extent, that element is instrumental in assessing the record's authenticity. On a more basic level, the Template serves as a checklist with definitions that help users determine whether they actually are even dealing with a record. Another very practical InterPARES 1 tool is what is informally referred to as the **Authenticity Requirements**. [8] This tool consists of two sets of requirements for assessing and maintaining the authenticity of digital records, with one set for records creators and one set for records preservers. The former set, known as the *Benchmark Requirements*, constitutes the requirements that support the presumption of the authenticity of a creator's digital records before those records are transferred to the custody of the preserver. The latter set, known as the *Baseline Requirements*, consists of the requirements that support the production of *authentic copies* of digital records transferred to the custody of the preserver and maintained within the preserver's preservation system. [9]

Perhaps the most fundamental of the InterPARES 1 conclusions is that in the digital environment, no original survives. In fact, the research concluded that it is not possible to preserve a digital record; it is only possible to preserve the ability to reproduce a digital record by processing the record's *digital components.* [10] For this reason, every successful processing of the digital components by the creator that results in a faithful copy of a record's content and of its documentary form is to be considered a *copy in form of original*, which, as the most reliable type of copy, is equivalent to the original as to its consequences, but generated subsequently. [11] Once in the custody of the preserver, an analogous process is used to generate authentic copies of the last instantiation of the creator's records. However, ensuring that what the preserver generates are indeed authentic copies requires continuous assessment and maintenance of the authenticity of the records throughout their lifecycle.

Key tools developed by InterPARES 2 include: (1) the **Framework of Principles**, comprising two complementary sets of principles for the creation and preservation of authentic digital

records, which together help structure the relationship between records creators and preservers by providing guidance for establishing a comprehensive intellectual framework within which creators and preservers can develop consistent and integrated policy environments conducive to effective and coordinated digital records preservation; (2) the **Guidelines for Preservers**, which provides concrete advice to any organization responsible for the long-term preservation of digital records; (3) the **Creation and Maintenance Guidelines**, which provides practical advice to individuals and small organizations for creation and maintenance of authentic digital content objects, including records, spanning in scope from selection of hardware and software to provisions for long-term preservation; (4) two comprehensive records management models: the **Chain of Preservation (COP) Model**, which adopts the perspective of the preserver, and the **Business-driven Recordkeeping (BDR) Model**, which adopts the perspective of the creator. [12] These two models depict, in both graphical and narrative form, all the activities and important, specific actions that must be undertaken, together with their inputs, outputs, constraints or controls and enabling mechanisms, to create, manage and preserve reliable and authentic digital records. As well, both models characterize the data and information that must be gathered, stored, and utilized to support the various management processes throughout the life of a record. The COP Model, which is based on the traditional 'records lifecycle' [13] approach, adopts the perspective of the records preserver (i.e., archivist or trusted custodian) "looking into the 'business' of a creating organization and identifying the records that are deemed necessary to preserve for internal business needs, or are likely to contribute to wider historical or societal objectives and interests." [14] In contrast, the BDR Model, which is based on the 'records continuum'[15] approach, adopts the perspective of the records creator "addressing its own 'business' within broader juridical, economic, and cultural contexts, and the records generated by that business. The viewpoint includes both those records needed for current business and those that need to be retained and preserved for the longer term historical interests of society;" [16] and (5) an online **Metadata Schema Registry**. This registry, officially dubbed the Metadata and Archival Description Registry and Analysis System (or **MADRAS**), is a centralized repository of schemas intended to aid in the identification of metadata sets, or the combinations of elements from different sets, that are appropriate to serve various recordkeeping and long-term preservation needs. The registry provides recommendations for how each schema might be extended or otherwise revised to address the reliability, authenticity and preservation needs of digital records created within the domain, community or sector to which they pertain. Currently in beta version, MADRAS is scheduled for official release to the general public in June 2007. [17]

The Concept of Authenticity

InterPARES 1 defined *authenticity* as "the quality of being authentic, or entitled to acceptance; as being authoritative or duly authorized, as being what it professes in origin or authorship, as being genuine," [18] while further clarifying that, in common usage, "*authentic* means 'worthy of acceptance or belief as conforming to or based on fact' and is synonymous with the terms genuine and bona fide," where "*genuine* 'implies actual character not counterfeited, imitated, or adulterated [and] connotes definite origin from a source' [and] "*bona fide* 'implies good faith and sincerity of intention.'" [19] Thus, with respect to records in particular, it follows that authenticity refers to "the trustworthiness of a record as a record;" that is to say, "the quality of a record that is what it purports to be and that is free from tampering or corruption." [20] Traditionally, in both archival theory and jurisprudence, assessment of authenticity of records that the creator relies on in the usual and ordinary course of affairs has typically been carried out by inference; in other words, with paper records, the authenticity of the record has been presumed unless proven to the contrary. However, because of the relative ease with which digital records can be altered, either inadvertently or intentionally, it is now the case that the presumption of authenticity of a digital record must be supported by evidence that the record is what it purports to be *and* that it has not been modified or corrupted in essential respects since the moment it was created. [21] In many contexts, most notably legal and professional ones, simply asserting that a record is authentic is useless if that assertion cannot be proven. Doing so requires establishing the record's *identity* and demonstrating its *integrity*. *Identity* refers to the attributes of a record that uniquely characterize it and distinguish it from all other records, such as the names of the persons concurring in its creation (i.e., author, addressee, writer and originator), its dates of creation and transmission, its relationships with other records, an indication of the action or matter to which it pertains, etc. [22] In effect, the identity attributes are the bare minimum metadata that must always be preserved with the record. *Integrity*, on the other hand, defined by InterPARES 2 as "the quality of being complete and unaltered in all essential respects," [23] refers to the wholeness and soundness of a record, which, when dealing with digital records, is assessed using integrity metadata.

InterPARES 2 Policy Cross-domain

The primary task of the Policy Cross-domain, one of four cross-domains in the InterPARES 2 Project (Figure 1), [24] was to identify and examine the policies and strategies that impact, influence or otherwise create barriers to the preservation of authentic digital records produced in the course of artistic, scientific, and e-government activities. This was a daunting task, made all the more challenging by the fact that records creators in all three Focus areas continue to adopt and rely on increasingly rich yet dynamic and thus somewhat unstable technologies without adequately considering, let alone resolving, the preservation challenges that these new technologies generate. Such activity underpins and is, to a large degree, directly responsible for perpetuating and exacerbating the challenges associated with developing effective strategies for the long-term preservation of authentic digital records. As the research of the Policy Cross-domain subsequently determined, the negative impact of these activities on preservation efforts is further compounded by the realization that:

> New models for collaboration and production, the outsourcing of activities and functions, and the privatization of many parts of the public domain, introduce new challenges for records retention. Legislation, case law, and multi-national agreements form an intricate and

> often inconsistent and internally conflicting regulating infrastructure that, rather than facilitating the proper creation and use of digital entities, makes it increasingly complex.[4] Taken together, recent changes in technology, public policy, and business models have put at risk the ability of organizations to undertake some of the activities necessary for the preservation of records. [25]

The records creation environments that emerged from the Project's case studies, [26] and the regulatory environments for records creation, maintenance, and preservation that emerged from the Policy Cross-domain's policy studies, generally exhibited little-to-no cohesive integration or consistency. With the noteworthy exception of those within the sphere of evidence law, very few of the organizations analyzed had in place a mature enough policy framework to even begin to address the digital preservation challenge. This deficiency was particularly evident in organizations whose activities involved complex, multi-component digital records. Based on these findings, in concert with the findings of the Project's other research teams and the case study reports, the Policy Cross-domain identified four principle policy themes or concerns, which, taken together, provide a good, high-level overview of the main issues at hand.

	FOCUS 1 Artistic activities	**FOCUS 2** Scientific activities	**FOCUS 3** Governmental activities
DOMAIN 1 Records creation and maintenance	Working Group 1.1	Working Group 1.2	Working Group 1.3
DOMAIN 2 Authenticity, accuracy and reliability	Working Group 2.1	Working Group 2.2	Working Group 2.3
DOMAIN 3 Methods of appraisal and preservation	Working Group 3.1	Working Group 3.2	Working Group 3.3
Terminology Cross-domain			
Policy Cross-domain			
Description Cross-domain			
Modeling Cross-domain			

Figure 1. *Matrix depicting Intellectual Framework of InterPARES 2 Project*

1. An inclusive policy infrastructure for record-keeping is required to support the activities of a society heavily reliant on information technology

Today's records creators and preservers are finding that their records management activities are being influenced and impacted by an increasingly complex landscape of legal, ethical and moral obligations, as well as community expectations (e.g., the use of records for accountability purposes). In response, many creators feel a need, and in some cases are in fact required, to introduce security, intellectual property and privacy rights management technologies to meet these obligations and expectations, all of which ultimately further compound long-term preservation efforts. The impact of the adoption of these access and redistribution control technologies (also known as digital rights management or DRM) on digital preservation efforts is exacerbated by the increasing transfer of information across networked and inter-connected organizational boundaries. Indeed, the confluence of these two developments, together with, as noted earlier, the adoption of increasingly rich yet dynamic and thus somewhat unstable technologies without adequate forethought as to the consequences of these actions on the full spectrum of records management needs, has resulted in a far more complex, and, in many ways, less reliable and more tenuous records management environment than has hitherto existed, and which "render[s] the already considerable challenge of preserving digital records far more complex than simply overcoming issues of technological obsolescence." [27] In short, the very same technological and legislative features that serve to protect privacy and intellectual rights, while enhancing immediate access to records and information for qualified users, invariably impede the ability of preservers to maintain these records for their "second non-commercial life." [28] For example, in cases where preservers are not able to secure exemptions from liability under intellectual property or privacy rights, they may be forced to anonymize records, thus compromising the integrity (and hence, authenticity) of the records. In general, preservation actions that result in changes to records at the bit level may be acceptable, while those that result to changes at a functional level likely will not. However, preservers must also bear in mind that *any* preservation action that alters a record, even at the bit level, could potentially contravene intellectual property rights. [29] Likewise, preservers may find their preservation efforts severely constrained, if not effectively thwarted, by moral obligations not to change the digital creation of an artist, where, for example, the creation relies on short-lived technological components. [30] Where residual rights or obligations (e.g., privacy, intellectual property, security, etc.) subsist within records, and where normal preservation actions would run the risk of contravening these rights and obligations, the preservation process assessment should incorporate a risk assessment component. [31] In other words, the preserver, who wishes to maintain authentic copies of the creator's records,

> must, in effect, be guided by the same concerns as the creator. That is, if the creator had to observe requirements of privacy, intellectual property, and security while maintaining the records, the preserver must also observe those requirements within the preservation environment, unless explicitly exempted. The foremost principle that must guide the long-term preservation of digital records was established in the first phase of InterPARES, which is to ensure that through preservation processes, records remain authentic copies of the creator's records. [32]

Finally, this reality highlights another key policy aspect that preservers must consider during their assessments; namely, the nature of their relationship with the creator(s). The relevance of this issue was first articulated by InterPARES 1 in relation to the

vital need during the records appraisal process for *determining the feasibility of preservation*, which is a task of the preserver that involves identifying "the elements and digital components of the records being appraised, and reconcil[ing] their preservation requirements with the preserver's current and anticipated preservation capabilities, and providing documentation about the digital components to be preserved and the feasibility of preservation." [33] Moreover, because there may, over time, be changes in the way a creator generates or organizes its records, or in the technology used to create them, it is important that the preserver monitor such activities of the creator in relation to the most current preservation feasibility assessments so that these assessments can be updated, as necessary. Obviously, the practicability of such a process hinges on the ability of the preserver to establish a clear, sustained, reciprocal relationship with the creator(s) whose records the preserver eventually intends to acquire, where both creator and preserver share an obligation to inform each other of procedural variations to records management activities and changes to technological practices and capabilities that may adversely impact the long-term preservation regime.

Consequently, clear and explicit acknowledgement of the above concerns, their inter-relationships and their impact on the management and use of the records in the custody of the preserver is absolutely vital for assessing the efficacy of an archives' preservation (or, in the case of records creators, preservation-friendly or preservation-enabling) processes and for developing an inclusive policy infrastructure in response to such an assessment that will support effective records preservation strategies. Clearly, however, the degree to which these sorts of issues must be addressed by the preserver during the preservation process assessment will depend on various situation-specific factors, including the nature of the legal and regulatory environment in which the creator(s) and preserver operate, the nature of the access and redistribution control technologies to which the records have been subjected, the nature of the creator(s) (and, hence, the records) in question, the nature of the preserver's relationship to the creator(s), etc.

2. An expanded and more detailed definition of record is necessary

As a consequence of the rapidly increasing adoption (by records creators) of technologies designed to facilitate both the transfer and virtual integration of information across networked and interconnected organizational boundaries, there has been an attendant increase in the number of different types of digital content objects—including some very complex, multi-component objects—that are now being created, many of which are already, or else have the potential of, being treated as records by their creators. This situation raises two important challenges with respect to preservation process assessments. First, preservers must be prepared to continuously re-evaluate their understanding of what documents each creator treats as records (i.e., those that the creator relies upon in its usual and ordinary course of affairs, associates with other records participating in the same activity or function, and refers to as the records of its affairs), and ensure that these documents are included in the assessments. This treatment is more consistent with the inclusive definition of "record" that is codified in most statutes. Second, preservers must be cognizant of a new category of records, identified by InterPARES 2 as "potential" or "prospective records," that are emerging as a consequence of the increased use of interactive and dynamic systems for augmenting information management, decision-making and records creation. This new category of records is summarized by the Policy Cross-domain as follows:

> Records have traditionally been identified as such retrospectively, that is, after having been completed and issued with a fixed form and stable content: but, with dynamic systems, there is the possibility of identifying "prospective" records. The entities that clearly manifest themselves as records since the moment they are created fulfil the traditional, memorial function of records to bear witness to or remember an action in which they participated or of which they were the residue. Rather than witnessing the past, prospective records guide the future through a set of instructions or actions to be carried out.[15] As such prospective records may not be considered records when their process of development begins, but, since their content can be fixed and their documentary form and functionalities described to make it possible to recreate them in the future, they could become records. Establishing policies to manage record-keeping for entities that are prospective records and *may* become records appears to fall into the context of guides, manuals, and other directive or procedural documents. [34]

Acceptance by preservers of a new conceptual understanding of the nature of the record that is extensible to the 'new' and rapidly evolving records management environment, and its use together with the related policy principles to ensure that important records are not overlooked during preservation process assessments, must be balanced against the danger of encouraging concepts of records that are *too inclusive* or, even more worrisome, *inconsistent*. Inconsistent and overly inclusive definitions of record can seriously compromise an organization's ability to comply with relevant statutes and to correctly interpret precedents set out in court decisions regarding records. [35]

3. Business processes are divided between many systems

As was demonstrated by many of the Project's case studies, the adoption by creators of increasingly rich and complex technological systems to create, capture and manage data is, at present, typically undertaken without due consideration of the functional attributes and limitations of these systems, especially in relation to their complex and sometimes distributed and distributing nature [36], and the impact that these characteristics have on the ability of such systems to adequately support the functions of records (i.e., as retrospective memorials or residues of action retained by their creator for reference or use in subsequent activities, or as prospective instruments of direction or instruction for future activities), and, ultimately, the long-term preservation of the records associated with such systems.

The growing practice among records creators of sub-dividing a business process between systems, or system components, each

with potentially varying degrees of complexity and dynamicity, led the Policy Cross-domain to suggest a need for policy direction among creators and preservers that is "as comprehensive as the systems and business processes at hand...[in which] records identified in one system [are] considered along with records related to the same business process created by other system(s) to ensure the most effective management, disposition, and preservation of records takes place." [37] More specifically,

> Policy should ensure that 1) the identification of documentary entities, including but not limited to records/metadata/ linkages, etc.,[17] is undertaken at the system design phase, 2) appropriate functions are incorporated to manage and preserve the entities identified at the outset of system development, and 3) the process and outcomes of these activities are reviewed regularly as part of system operations. [38]

Identification of the presence/absence, and/or evaluation of the effectiveness of the implementation of these recommendations, should be incorporated into preservation process assessments so that creators and preservers can properly address and monitor the potential impact that the atomizing records management practices noted above have on the creator's ability to create and maintain authentic and preserveable records, and the preserver's ability to identify, appraise, acquire and retain authentic copies of the creator's records for the long term. Assessors are encouraged to make use of the other aforementioned InterPARES 1 and 2 tools to help pinpoint more precisely where and in what ways these recommendations may manifest themselves in relation to a particular creator's unique system and records management activities, and how these, in turn, may relate to the preserver's own system and preservation requirements.

4. Preservation policies are inadequate or absent

The records considered worthy of preservation may not be preservable because of "quick-fix" records management decisions that have not been adequately considered with respect to their impact on long-term preservation, such as the use of encryption or digital signatures, for example. Moreover, as was observed in the case studies, records creation and management activities are very often directed at safe-guarding data, not records. Individually, the potential impact of these practices is serious enough on long-term preservation efforts; when combined, however, the detrimental impact is likely to be pervasive and irreversible. For example, although back-up and disaster recovery routines were found to be widespread, the sophistication of these routines was, in the vast majority of cases, extremely rudimentary and/or ad hoc (e.g., consisting exclusively of burning data to CD-ROMs and supported by no established and documented policies and procedures), with no consideration given to such a vital concern as interoperability across time (i.e., so that records originally stored in one system can in fact be restored to a subsequently upgraded or otherwise modified system). As is perhaps most succinctly and effectively demonstrated by the Project's two records management models, creators and preservers need to better coordinate their activities so that more inclusive, compatible and effective preservation policies can be developed that address the preservation of records across their entire lifecycle (or continuum), regardless of their location in the chain of custody at any point in time, or the nature of the system in which they reside. Again, this is another area of concern that should be incorporated into preservation process assessments, preferably augmented through reference to the aforementioned models.

Conclusion

It is clear from the foregoing that digital preservation extends well beyond the early basic concerns of technological obsolescence and media fragility, to include a multitude of rapidly evolving and confounding non-technological preservation concerns related to, among other things, various organizational process, policy and socio-cultural issues. Moreover, it is clear that digital preservation is, fundamentally, a *distributed* process involving "a range of different (and often differently interested) stakeholders who become involved with digital resources at particular phases of their life cycle." [39] Consequently, creators and preservers must develop new strategies for instituting and sustaining more active and integrated records management collaborations that take into account authenticity concerns throughout the entire chain of custody, from creation through preservation. Ideally, these collaborations should be informed and supported by a comprehensive and *harmonized* intellectual framework of policies, procedures, practices, and standards. Preservation process assessments, informed by model frameworks such as the one provided in the InterPARES *Framework of Principles*, are an important tool for helping creators and preservers identify and isolate the key *deficiencies* in their current frameworks, especially with respect to irreconcilable or conflicting records management activities in relation to the full records lifecycle or continuum, so that the necessary steps can be taken by both creators and preservers to better harmonize their activities in support of the long-term preservation of authentic digital records.

Although this paper has focused primarily on the *Framework of Principles*, various other tools developed by InterPARES 1 and 2 provide comprehensive guidance for all aspects of managing digital records, from creation through preservation, and thus can be used by creators and preservers to augment their preservation process assessments. These other tools include: (1) the *Template for Analysis*; (2) the benchmark and baseline *Authenticity Requirements*; (3) the *Guidelines for Preservers*; (4) the *Creation and Maintenance Guidelines*; (5) two comprehensive records management models: the *Chain of Preservation Model*, and the *Business-driven Record-keeping Model*; and (6) the *Metadata and Archival Description Registry and Analysis System* (MADRAS).

References

[1] K. E. K. Brown, "Use of General Preservation Assessments: Process," Libr. Resour. Tech. Serv. 49(2), 90 (2005). See also references cited therein.

[2] Two pioneer studies in this area include: M. Hedstrom and S. Montgomery, "Digital Preservation Needs and Requirements in RLG Member Institutions," (Mountain View, CA, Research Libraries Group, 1999). http://www.rlg.org/preserv/digpres.pdf; and D. Lievesley and S. Jones, "An Investigation into the Digital Preservation Needs of Universities and Research Funders: the Future of Unpublished Research Materials ," British Library RIC Report no.

109 (1998). http://www.ukoln.ac.uk/services/papers/bl/blri109/; More recent examples include: M. Jones and N. Semple, "Mind the Gap: Digital Preservation Needs in the UK," Ariadne 48 (July 2006). http://www.ariadne.ac.uk/issue48/semple-jones/ ; D. Simpson, "Digital Preservation in the Regions: Sample Survey of Digital Preservation Preparedness and Needs of Organizations at Local and Regional Levels," (Museum, Libraries and Archives Council, London, 2005). http://www.mla.gov.uk/resources/assets//M/mla_dpc_survey_pdf_6636.pdf ; and T. Clareson, "NEDCC Survey and Colloquium Explore Digitization and Digital Preservation Policies and Practices," RLG DigiNews 10(1), (2006). http://www.rlg.org/en/page.php?Page_ID=20894#article1.

[3] C. T. Cullen, "Authentication of Digital Objects: Lessons from a Historian's Research," in Authenticity in a Digital Environment, CLIR report 92 (Council on Library and Information Resources, Washington, D.C., 2000), pg. 3. http://www.clir.org/pubs/abstract/pub92abst.html.

[4] L. Carpenter , "Supporting Digital Preservation and Asset Management in Institutions," Ariadne 43 (April, 2005). http://www.ariadne.ac.uk/issue43/carpenter/.

[5] Cullen, "Authentication of Digital Objects," op. cit.

[6] See http://www.interpares.org/.

[7] Authenticity Task Force, "Appendix 1: Template for Analysis," in L. Duranti, ed., The Long-term Preservation of Authentic Electronic Records: Findings of the InterPARES Project (Archilab, San Miniato, Italy, 2005), pg. 192-203. Also online at http://www.interpares.org/book/interpares_book_j_app01.pdf.

[8] Authenticity Task Force, "Appendix 2: Requirements for Assessing and Maintaining the Authenticity of Electronic Records," in L. Duranti, ed., The Long-term Preservation of Authentic Electronic Records: Findings of the InterPARES Project (Archilab, San Miniato, Italy, 2005), pg. 204-219. Also online at http://www.interpares.org/book/interpares_book_k_app02.pdf.

[9] An authentic copy is "a copy certified by an official authorized to execute such a function, so as to render it legally admissible in court" (InterPARES 2 Terminology Database, Glossary, s.v. "authentic copy." http://www.interpares.org/ip2/ip2_terminology_db.cfm). Note that, as discussed here, the production of authentic copies occurs in relation to two distinct, yet related, contexts: records transfer and records preservation. Regarding records transfer, this relates to the process by which the preserver (i.e., designated trusted custodian), after having first assessed the authenticity of the creator's records slated for transfer to the archives for long-term preservation, produces authentic copies of them from the creator's recordkeeping system. Once in the custody of the preserver, the authentic copies are then maintained by the trusted custodian in a trusted preservation system, which, by definition, must have in place explicit rules and procedures for the ongoing production of authentic copies that is able to account for such potentially confounding factors as system obsolescence, media deterioration, changes in technology, etc. It is important to emphasize, however, that reproduction of records in the preserver's preservation system does not ipso facto result in authentic copies; instead, such designation must explicitly be provided by the preserver's authority.

[10] A digital component is "a digital object that is part of one or more digital records, including any metadata necessary to order, structure, or manifest the content, requiring a given preservation action," where a digital object is understood to mean "a unit of digital information that includes properties of the object and may also include methods of performing operations on the object" (InterPARES 2 Terminology Database, Glossary, s.v. "digital component" and "digital object." http://www.interpares.org/ip2/ip2_terminology_db.cfm).

[11] InterPARES 2 Terminology Database, Glossary, s.v. "copy in form of original." http://www.interpares.org/ip2/ip2_terminology_db2.cfm

[12] See http://www.interpares.org/ip2/ip2_models.cfm.

[13] "Records lifecycle" is defined as "the theory that records go through four distinct stages of change in activity, including creation or receipt, use and maintenance, in-active storage, and disposition (destruction or transfer to an archives) (InterPARES 2 Terminology Database, Dictionary, s.v. "records lifecycle." http://www.interpares.org/ip2/ip2_terminology_db.cfm. For a multi-faceted introduction to the lifecycle concept, see: Proceedings of the DLM-Forum on Electronic Records, Brussels, 18-20 December 1996 (European Communities, Luxembourg, 1997). http://europa.eu.int/ISPO/dlm/dlm96/Proceed_en.htm.

[14] T. Eastwood and H. Hofman, Draft Final Report of the Modeling Cross-domain (InterPARES 2 Project, Vancouver, 2006), pg. 6.

[15] "Records continuum" is defined as "the whole extent of a record's existence...[and] refers to a consistent and coherent regime of management processes from the time of the creation of records (and before creation, in the design of recordkeeping systems) through to the preservation and use of records as archives" (Adapted from Standards Australia, AS 4390, Part 1, Clause 4.22). For a comprehensive summary of the records continuum concept, see X. An, "An Integrated Approach to Records Management," Info. Mgmt. J., 37(4), 23-32 (2003). See also Monash University's Records Continuum Research Group Web site at: http://www.sims.monash.edu.au/research/rcrg/.

[16] Eastwood and Hofman, Modeling Cross-domain, op. cit.

[17] See http://www.gseis.ucla.edu/us-interpares/madras/.

[18] InterPARES 1 Project, "The InterPARES Glossary," in L. Duranti, ed., The Long-term Preservation of Authentic Electronic Records: Findings of the InterPARES Project (Archilab, San Miniato, Italy, 2005), pg. 357. Also online at http://www.interpares.org/book/interpares_book_q_gloss.pdf.

[19] H. MacNeil et al., "Part One, Establishing and Maintaining Trust in Electronic Records: Authenticity Task Force Report," in L. Duranti, ed., The Long-term Preservation of Authentic Electronic Records: Findings of the InterPARES Project (Archilab, San Miniato, Italy, 2005), pg. 21. Also online at http://www.interpares.org/book/interpares_book_d_part1.pdf. Emphasis as in original.

[20] InterPARES 2 Terminology Database, Glossary, s.v. "authenticity." http://www.interpares.org/ip2/ip2_terminology_db.cfm.

[21] MacNeil et al., "Authenticity Task Force Report," op. cit.

[22] Authenticity Task Force, "Appendix 2," op. cit., pg. 205.

[23] InterPARES 2 Terminology Database, Glossary, s.v. "integrity." http://www.interpares.org/ip2/ip2_terminology_db.cfm. As previously clarified by InterPARES 1, the notion of 'in all essential respects' "does not mean that the record must be precisely the same as it was when first created for its integrity to exist and be demonstrated" (MacNeil et al., "Authenticity Task Force Report," op. cit., pg. 47). Indeed, although InterPARES 1 established fixed form

and stable content as two of the essential characteristics of a digital record, it is important to recognize that these constraints are not absolute. As in the paper world, where, with the passage of time, records are subject to deterioration, alteration and/or loss, digital records, because of the fragility of the media, the obsolescence of technology, the idiosyncrasies of electronic systems, and trauma from intentional or accidental mishandling, are likewise susceptible to loss or corruption of elements of form or content. However, as long as such factors do not compromise the essence of the record – that is to say, the message that it is meant to communicate to achieve its purpose – then it remains a trustworthy record. In other words, "this implies that [a digital record's] physical integrity, such as the proper number of bit strings, may be compromised, provided that the articulation of the content and any required elements of form remain the same." (Ibid.) InterPARES 2 researchers further expounded on this issue with the introduction of the concept of "bounded variability," which refers to the notion that changes to the form and/or content of a digital record that are limited and controlled by fixed rules (due either to the inherent functionality of the system in which the record resides or to the intention of the author or writer of the document) do not violate the requirements of fixed form and stable content (see L. Duranti and K. Thibodeau, "The Concept of Record in Interactive, Experiential and Dynamic Environments: the View of InterPARES," Archival Science 6, 13-68 (2006).

[24] The task of the cross-domains was to address research questions within each cross-domain's purview that were common to all areas of inquiry in the Project.

[25] L. Duranti, J. Suderman and M. Todd, Policy Cross-domain Final Report (InterPARES 2 Project, Vancouver, 2007), pg. 4. http://www.interpares.org/ip2/ip2_documents.cfm?cat=policy. Note: footnote reference in the quote is from the original text, and is not reproduced here.

[26] Throughout the duration of the InterPARES 2 research, the three focuses completed 22 records creator-based case studies. The data from these case studies, together with the diplomatic analysis and modeling activities carried out on them, are the core research data of InterPARES 2. For a detailed summary and comparative analysis of the case studies and the data they generated, see M. Cardin, "Domain 1: Records Creation and Maintenance, Final Report," (InterPARES 2 Project, Vancouver, 2007). http://www.interpares.org/ip2/ip2_domain1.cfm.

[27] Duranti et al., Policy Cross-domain Final Report, op. cit., pg. 20.

[28] Ibid., pg. 19. As cited in Duranti et al., "'Second noncommercial life' has been elaborated by legal scholar Lawrence Lessig as the period that begins when the copyright term expires and content becomes subject to re-use; see Lessig's Free Culture (The Penguin Press, 2004), and the Editorial, "The Coming of Copyright Perpetuity," New York Times, Jan. 16, 2003, p. A28."

[29] Ibid.

[30] Ibid., pg. 22.

[31] Ibid., pg. 23.

[32] Ibid., pg. 19.

[33] InterPARES 2 Terminology Database, Glossary, s.v. "determine feasibility of preservation" http://www.interpares.org/ip2/ip2_terminology_db.cfm. As outlined by the Modeling Team, "the details of this process are 1) to identify the necessary documentary components (e.g., record profile, attachments, annotations, etc.) and elements of form (e.g., author, date, subject line, etc.) of records to be preserved to determine which record elements must be preserved to protect the authenticity of those records; then 2) to identify the digital components that manifest the record elements that need to be preserved to protect the authenticity of records earmarked for permanent preservation; and, finally, 3) to determine whether the digital components manifesting the record elements that need to be preserved to protect the authenticity of records earmarked for permanent preservation can in fact be preserved given the preserver's current and anticipated preservation capabilities" (Eastwood and Hofman, Modeling Cross-domain, op. cit., pg. 41).

[34] Ibid., pg. 11-12. Emphasis as in original. For a more in-depth discussion of potential or prospective records, see Duranti and Thibodeau, "The Concept of Record," op. cit. Note: footnote reference in the quote is from the original text, and is not reproduced here.

[35] Duranti et al., Policy Cross-domain Final Report, op. cit., pg. 15-16.

[36] Distributed in the sense that individual components or functions of the system may be spread across a wide network of interconnected, yet physically and even administratively disparate units of an organization, and distributing in the sense that documentary elements that convey the semantic of a record, such as metadata schemas, may be stored elsewhere in the system as entities separate from the record (Ibid., pg. 12).

[37] Ibid., pg. 12-13.

[38] Ibid., pg. 13. Note: footnote referenced in the quote is from the original text, and reads as follows: "Because semantic value can be derived from an understanding of how documentary entities relate to one another (for example, a registry to a series of records and the records themselves), additional entities of interest might include data and system models, domain-specific taxonomies, and enterprise architecture models and specifications."

[39] N. Beagrie and D. Greenstein, A Strategic Policy Framework for Creating and Preserving Digital Collections, version 5.0 (Arts and Humanities Data Service, United Kingdom, 2001), pg. 3. http://ahds.ac.uk/strategic.pdf.

Author Biography

Randy Preston received his MA in archival studies from the University of British Columbia (2006), during which time he participated in the InterPARES (International Research on Permanent Authentic Records in Electronic Systems) 2 Project as a graduate research assistant for Case Study 14, General Study 09 and the Modeling Cross-domain. He is currently the InterPARES 2 Project Co-ordinator.