# International Research on Permanent Authentic Records in Electronic Systems (InterPARES) 2: Experiential, Interactive and Dynamic Records

This book presents the consolidated findings of the second InterPARES research project1. The InterPARES projects have examined issues associated with the creation, maintenance and preservation of digital records, primarily seeking to identify the means by which authentic and reliable digital records can be maintained through time. InterPARES 1 assessed these issues from the perspective of the record preserver while InterPARES 2 has taken the perspective of the records creator.

This change is significant. The first InterPARES project was firmly based in life cycle theory but InterPARES 2, taking its lead from the research findings of the first project which found, in part, that preservation activities designed to maintain authentic records could not be deferred until the traditional latter stages of the life cycle model, sought also to embrace aspects of continuum theory. The success or otherwise of this approach will be discussed further below.

The scope of the InterPARES 2 research project as represented by this publication is breathtaking. The project ran from January 2002 to December 2006. The project team was highly interdisciplinary, involving researchers from a broad range of artistic, scientific, legal and information based professions. The team was also international, composed of researchers from Africa, Europe, North America, Asia and Australia, with many of those participating rightly regarded as luminaries in the areas of record keeping and archival research. The project's objectives were ambitious, seeking to understand records generated by "experiential, interactive and dynamic systems", with particular reference to record keeping practices in the artistic, scientific and government sectors, and attempting to determine the concepts, principles, criteria and methods by which reliable and authentic records could be created and preserved within these sectors.

The introduction to the book outlines the complex structure of the research project. The base data used for analysis within the project was a series of case and general studies of a diverse range of artistic, scientific and governmental projects. The case studies were used to examine the records and record keeping practices of a specific records creator whilst the general studies were used to conduct broader research into a specific issue of interest to the project team, such as the development of data portals in the sciences (General study 10), the use of digital technologies by photographers (General study 7) or the functionality of government websites (General study 8).

The data produced from these 32 case and general studies was fed into three different research areas known as "domains" within the broader project team: Records Creation and Maintenance (Domain 1); Authenticity, Accuracy and Reliability (Domain 2); and Methods of Appraisal and Preservation (Domain 3). In addition, there were four cross-domain working groups – Modelling, Description, Policy and Terminology – that sought to address policy, description, modelling and terminology issues across the creation, maintenance and preservation divides. Each domain or cross-domain assessed the case and group study data to determine answers to their own specific research questions. The final reports of each of the domains or cross-domains represent the seven parts of this book.

This then is the structure of the book and the structure of the research project upon which it is based. So what about the content? Has this mammoth research project generated findings that will be useful to those struggling with digital record keeping issues? Does it provide actual guidance as to how authentic and reliable digital records can be created and

maintained?

The answer is yes, but because of the size of the publication, you do have to work hard to find it. For practitioners coming to this work, I would recommend the *Framework of Principles for the Development of Policies, Strategies and Standards for the Long-term Preservation of Digital Records* (Appendix 19), as well as the *Creator Guidelines Making and Maintaining Digital Materials: Guidelines for Individuals* (Appendix 20) and *Preserver Guidelines Preserving Digital Records: Guidelines for Organizations* (Appendix 21) as useful and immediately accessible consolidated research products. These publications distil significant amounts of the iterative scholarly research contained throughout the book into clear, simple guidance for a general audience about safeguarding the long-term accessibility of digital records. The *Glossary* produced by the Project (found at the end of this book) is voluminous and provides a large number of definitions for many of the technical requirements associated with digital records preservation and this also may be of immediate practical benefit to practitioners.

The rest of the project findings contained in this publication are not as immediately accessible. The detail contained in the book can be quite overwhelming and the Project's website is itself not easy to navigate to find clear and comprehensive deliverables. I therefore would have preferred more shorter, pithier research summaries similar to Creator Guidelines and Preserver Guidelines that focus on key metadata lessons learned, or key recommendations for the scientific community, for example, as an additional means to promote and disseminate research findings. Detailed recommendations on these and many other issues have been made throughout this book and in numerous other scholarly publications and conference papers associated with the research project but, in our age of the sound byte, simple, short summaries can have a significant impact. To some they may seem to simplify a message too significantly, hiding necessary complexities and "dumbing down" research findings. But to me the broader audience that a simplified message brings outweighs any possible negative consequences.

Given too the interdisciplinary focus of the research, it would have been useful if more of the project findings could be communicated in ways that cross disciplinary boundaries. This book is written in record keeping language for record keepers to read. At present, many of the research findings presented in the book would be too dense and complicated to have meaning for the artistic, scientific or even government communities that contributed to their development. Given the scope, extent, importance and timeliness of the InterPARES 2 research, I would like its findings to be disseminated as clearly and as broadly as possible so that its significant message can be heard. Governments, business leaders, technology innovators, artists, scientists, funding institutions and individuals in their own record keeping capacities all need to be aware of the fragility of the digital world and aware of the basic steps that will enable them to preserve authentic digital records. I would like to see the findings of the InterPARES research promoted more in ways that would make them accessible to these audiences. I believe this would be an appropriate end result for the incredible amount of deep scholarly research work that has gone into the development of this publication.

So pithier "executive summaries" would have been nice, but what of the remainder of the book as it currently stands? If you have the time, it is still well worth a read.

For me, Part One, the Focus Task Force Report which provides an overview of the artistic, scientific and government case and general studies compiled in the course of the research was a fascinating summary of a diverse range of record keeping environments and the problems that they face. Interestingly, for all their diversity, each of the record creators generally encountered the same set of problems – issues of technological obsolescence, problems with long term accessibility, lack of control over creation, problems with sustaining authentic and accurate data created in proprietary systems, insufficient metadata documentation and uncertainty as to which records actually need to be captured2. Part Four of the publication also notes the reluctance identified in the case studies of individuals and organisations to take responsibility for and action in response to preservation problems. For example, in case study 21, the Supreme Court of Singapore chose to outsource the management of the Public Key Infrastructure that supports the electronic submission of court records so that it can be "insulated from managing technology obsolescence"3. This naivety in relation to responsibilities for preservation and the extent of the preservation problem is a theme that recurs throughout the publication.

Part Three, the report of the Authenticity, Reliability and Accuracy Domain, teases out many interesting insights into the nature of authenticity, reliability and accuracy that the research team derived from assessment of the case and general studies. As one who works in the government sector, for me this section of the publication provided an interesting window into other areas of record keeping practice and the specific nature of the problems that they face. The fragility of art, for example, in the digital age is a pressing issue, with case study 13 in particular, through its attempts to migrate a musical piece originally composed on now obsolete hardware and software, demonstrating that the long term preservation of accurate and authentic digital works may not, in many instances, be possible. This section of the book also provided an interesting discussion of the different understanding of "the record" in the scientific professions. The project research revealed that often in the sciences the ultimate "record" of research is a published article in a scientific journal. The "data" that went into the compilation of that "record" is regarded as more ephemeral and of less interest to fellow researchers than the publication. The data is also frequently regarded as likely to be supplanted by other scientists' findings in years to come. The vexed issue of what is a record is therefore still alive and well and does need to be reconsidered and potentially redefined in the digital age.

Luciana Duranti and Kenneth Thibodeau do conduct this form of reappraisal in Appendix 2 to the book, "The Concept of Record in Interactive, Experiential and Dynamic Environments: the View of InterPARES", a reprint of their 2006 *Archival Science* article. This Appendix revisits the very traditional understanding of "record" that underpinned the InterPARES 1 research project and shows how the experience of assessing dynamic business systems forced the project team to redefine the diplomatic concept of a record so that it can accept the outputs of highly dynamic business systems as records. According to traditional diplomatic definitions, records must have recognisable documentary form, but record form is not immediately apparent in dynamic business applications. Project research demonstrated that by identifying relevant metadata and business rules a fixed representation can be made of transactions conducted in such systems. Therefore these systems can produce and maintain identifiable records, albeit in ways and means that require a reconceptualisation of traditional diplomatic understandings. This reconceptualising of the record has been carried through into Part Four of the book in its discussion of appraisal and preservation.

Throughout these and other areas of the book there is a strong emphasis on the critical role metadata plays in digital preservation. Part Three highlights the importance of metadata to presumptions of authenticity in the scientific community. Among scientists "it is considered axiomatic that a database has limited utility unless the auxiliary information required to understand and use it correctly – the metadata – is included in the record"[4]. In Part Six of the book the Description team examines current metadata practice as represented by the case study data and reveals that, with the exception of a few scientific and government bodies, there was little attempt in the case study organisations to implement metadata at all, in spite of its critical importance for digital preservation and authentication. From this the Description team then extrapolates ways and means of improving current metadata application.

A key metadata product discussed in Part Six for improving metadata performance is MADRAS, the Metadata and Archival Description Registry and Analysis System. The genesis, development and testing of MADRAS is examined in detail in this section of the publication. MADRAS was designed to analyse the record keeping capabilities of metadata sets and to identify the specific areas where metadata schemas fail to address the record keeping requirements necessary for the preservation of authentic and useable records. MADRAS is a very comprehensive and well designed tool that should provide those looking to engage more with metadata an understanding of the capacities of metadata in relation to digital preservation and a knowledge of the existing metadata tools that can support their requirements. MADRAS is available online (relevant links and guidance for its use are provided in the book) both as an automated schema assessment tool and as a registry for identifying metadata schema and tracking their evolution through time.

Incredibly detailed and complex models of record processes that can also be used by practitioners to assess and improve their practices are provided in Part Five of the book. The Chain of Preservation model (described in Part Five and then presented in detail at Appendix 14) is based on the perspective of the organisation responsible for the long-term preservation of digital records and decomposes literally thousands of processes identified by the project team as comprising the means by which authentic records can be maintained through time. The Business Driven Record keeping Model takes the perspective of the records creating entity. It positions record keeping as an integrated part of organisational business, with business being defined here in the most inclusive sense.

It is in this section that the conceptual divides inherent in the project become most apparent. The Chain of Preservation (COP) model "proceeds from an understanding of the theory, methods and practice of proper records processes throughout the lifecycle of the records" and is ultimately "consonant with the perspective of the preserver"[5]. The Business-Driven Record keeping (BDR) Model aims "to identify and express the relationship between doing business activities and the information used and created in doing this, as well as the capture and management of that information as records"[6].

Where the COP model focuses very specifically on the record at all stages of its lifecycle, the BDR model looks at a record both in terms of its management requirements and also in terms of its relationships. Its relationships to business, to people, to the mandates that may prescribe requirements for its creation, use or management, its relationship to broader societal frameworks down to its relationship with very granular workflow processes. In these two models the differences between life cycle and continuum perspectives on records and record processes are writ large. The existence of the two models likely demonstrates that the research team was unable to reconcile the inherent differences between the lifecycle and continuum models and unable to achieve theoretical consensus on a preferred conceptual approach. There were obviously difficulties in integrating these theories and creating unified products from these two different world views[7]. Instead of integration, the two conceptual approaches have proceeded in parallel, as evidenced in this and other sections of the book.

Rather than a negative, however, I think this can be seen as a positive aspect of the research project. It demonstrates how far the project has come from its first phase, where record keeping concerns in the creation environment (a key component of continuum thinking) were not a research focus. It also demonstrates the truly interdisciplinary nature of the research team, and the team's willingness to accept and incorporate a range of theoretical viewpoints. The combination of research perspectives has also, in this section of the book in particular, created an incredible richness and depth of research where findings from one theoretical perspective can be used to inform the findings of the other[8]. One would hope that these incredibly rigorous and considered models would find a ready audience amongst software designers and business process engineers, as well as record

keeping professionals. There are plans in future phases of this research to integrate the research findings of the Description Domain with the BDR model to develop a metadata specification model (a metadata specification has already been developed for the COP model). When metadata specifications for both these models are complete, this could enable flexible, multi-entity, through time metadata descriptions that could then be mapped to specific processes to enable automated application of metadata requirements designed to support authenticity and ongoing accessibility. Closer integration with the conclusions of the Description Domain might also lead to the identification of certain standard processes or requirements that are necessary in all environments in order to yield authentic and robust digital records. If such specific requirements could then be communicated with software and hardware developers and disseminated through their technology, our digital heritage could be looking much more secure.

A key conclusion that can be inferred from the book is that improved record keeping structures to support digital preservation must start now. The News Archives Survey described in Part Six of the book makes this point clearly. This survey targeted contemporary news archives managing records in digital formats. The survey asked news librarians and archivists about their digital preservation practices. The findings make grim reading. Only 18 per cent of those surveyed had not experienced data loss during migration projects. (The losses identified ranged from minor – a few corrupt images on CD-ROMS – to disastrous – the complete loss of an entire collection of thousands of images)9. Only 33 per cent of those surveyed had management who were committed to preservation of the archives. The survey also demonstrated that the concept of maintaining authenticity through digital preservation was not well understood, with, for example, bit level authentication of individual files following preservation action virtually unknown amongst those responding to the survey. The survey also revealed that archivists and librarians in news archives are responsible for very large numbers of diverse file formats, generally employ no standard metadata schema, compile very little metadata about the digital dependencies of their records and have little or no budget for digital preservation. Taken together with the findings of the case and general studies presented through the publication, these studies show that digital preservation is a significant challenge facing us now that will have major ramifications on our ability to access critical data in the future.

For all its complexity then, the basic conclusions of this publication are simple and are ones that we all need to take heed of. Primarily we as record keeping professionals need to be aware of the critical issue of digital preservation and we then need to be passionate advocates of the importance of digital preservation within the broader community. The findings of both InterPARES projects demonstrate that digital records are all pervasive and planning for their preservation needs to be begin at or before their creation if, firstly, they are to be maintained and, secondly, if they are to be maintained as reliable and authentic records of action. The critical issue that comes through in the vast amounts of InterPARES research is that all communities need to be better informed of the vulnerabilities of digital information and then also made aware of the authentication and evidential requirements necessary to sustain records through time.

As noted in Appendix 19, "The case studies showed that record creation in the digital environment is almost never guided by considerations of preservation over the long term. As a result, the reliability, accuracy and authenticity of digital records can either not be established in the first place or not be demonstrated over periods of time … These records cannot therefore support the creator's accountability requirements, nor can they be effectively relied upon either by the creator for reference or late action or by external users as sources"10. This situation must be addressed if the preservation of authentic records is to be achieved.

For me, ultimately, the true value of this book is in its stories. Stories of artists creating multimedia performance pieces involving dance, music and video based interaction between performers. Stories of NASA scientists sending rockets to Mars and trying to document the data that is returned. Stories of Singaporean lawyers developing systems to allow for the legal submission of digital evidence. Stories of the variety of records that artists, scientists and government officials create and of the multitude of ways in which they manage them. These individual stories themselves are fascinating but together they have been used by the InterPARES team to create a sophisticated and comprehensive portrait of contemporary record keeping in its multitude of guises. The real world context provided by the case and general studies has both contributed to the development of and facilitated the testing of all of the Project's key deliverables – the principles, the COP and BDR models, MADRAS, the metadata specification models etc. There is a wealth of experience and insight in this publication for all with an interest in record keeping and digital preservation. I do wish that some of these insights were more accessible but this is a scholarly publication and is necessarily bound by the traditions of its form. Despite its size and complexity, the key message of the InterPARES 2 research is clear. Record keepers of the world unite. The future of our digital heritage needs you.

**Kate Cumming**
*State Records Authority of New South Wales, Kingswood, Australia*