DigiCULT.Info

CONTENTS

INTRODUCTION

This tenth issue of *DigiCULT.Info* presents a wide variety of articles and information on the latest developments and projects within the cultural heritage field in 2005. With a discussion of the integrity and authenticity of digital records, covered in our first Newsletter and Thematic Issue, and an update of the InterPARES project, DigiCULT has come a full circle. Luciana Duranti, of InterPARES 2 presents general and case studies which identify the issues and challenges of creating digital versions of subjective works, and describes the development of a Metadata Schema Registry, enabling users to select appropriate methods for their own record-keeping.

Records and information management continues to be one of the core interests of the cultural heritage sector. Document Management and Electronic Archiving (DOMEA), based in Germany, presents the administration of electronic records and how they can be used to support business processes, with examples of federal schemes and policies on the eventual disposal of electronic records. A technological perspective on digitisation and archiving is presented by the Department of Information and Knowledge Management of Budapest University of Technology and

© Ergani, 2005, http://www.ergani.org.gr
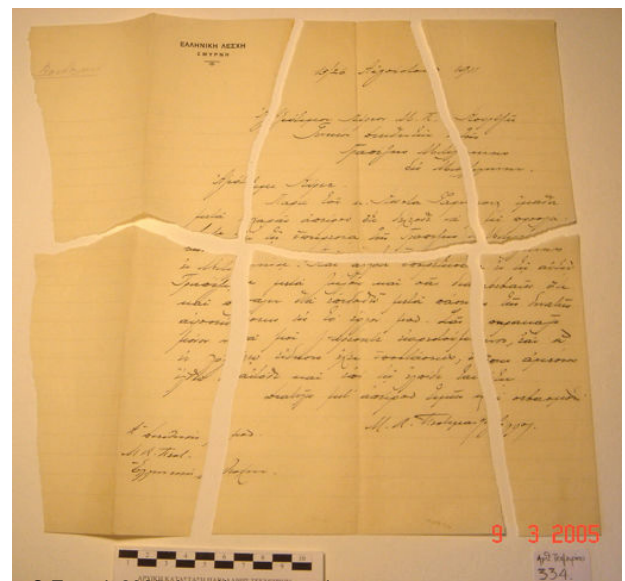
INTERPARES 2: A PROGRESS REPORT

Shaunna Moore and Luciana Duranti

In the very first issue of *DigiCULT.Info* in 2002, Luciana Duranti spoke to us about her work with InterPARES. For DigiCULT.Info's 10th issue, we asked for an update on this work.

Digital records only exist in the status of original at the moment when they are made or received. After a digital record has been saved for the first time, each retrieval produces a copy. These copies are still the creator's records, because they are used for action or reference in the usual and ordinary course of business of the person who makes, receives and accumulates them. However, they can only be considered reliable if the processes of creation and maintenance have been subject to strict controls and have respected authenticity requirements. When the creator's records are no longer active and are acquired for long-term preservation by some trusted custodian, we are confronted with a greater intellectual problem. In fact, it is not possible to preserve a digital record. It is only possible to preserve the ability to reproduce it as an authentic copy. Because copies made for purposes of preservation do not participate in the creator's course of business, they are no longer its records, although they may be endowed with the same reliability if the process of reproduction has been carried out by a competent person, following strict rules and properly documented.

Analysis of the relationship between authenticity and reproduction has always placed theorists in a precarious position. As the definition of authenticity includes both identity (the entity is what it purports to be) and integrity (the entity has not been corrupted), how can reproduction be undertaken in such a way that these qualities can be transferred from the original to the duplicate? Considering that a copy of an entity is never identical to its first instantiation, both facets of authenticity are immediately threatened upon attempts to imitate an original. As a result, the verification of authenticity in a reproduction necessitates a thorough understanding of the entity that is being reproduced in order to determine which specific parts, characteristics, attributes or formal elements are crucial to its identity, and to what extent alteration is acceptable.[57]

Nowhere is this more apparent than in the performing arts, which by their nature require the repeated presentation of interpretations of scores or scripts by different performers who execute the original idea in ever-new environments. In this context, the parameters of authenticity are unclear, for even in its original form a performance relies on a multitude of interactive components, which are never entirely determined by the creator of the work. The factors deemed relevant for the authentic reproduction of a performance will vary according to the circumstance. For instance, while some will consider the use of original musical instruments of the period to be of utmost significance in the performance of a baroque piece, others will focus on the articulation of notes, interpretations of dynamics, or even the ambience of the environment when evaluating whether the performance captured the essential quality of the composer's original score.

The concept of repetitive performance serves as a useful analogy in an electronic environment. The existence of entities that are fragmented into their digital components upon every storage action and reconfigured upon every viewing necessitates an in-depth analysis of the extent to which each reconfiguration may be deemed identical to the original entity. Clearly, this entails some degree of compromise with regard to the meaning of integrity, for, how can an object that has been torn apart and reconstructed truly be described as uncorrupted? The issue becomes further complicated by the highly interactive and dynamic nature of the modern digital environment. Authenticity in this case must take into account the ability to preserve some semblance of shifting contexts and relationships among and within digital objects.

INTERPARES 2

InterPARES (International research on Permanent Authentic Records in Electronic Systems, http://www.interpares.org/ip2_index.cfm) has been devoted to the development of the theoretical and methodological knowledge that will provide a framework for the long-term preservation of authentic records in electronic systems. As an international and interdisciplinary project,[58] the goal of the research is to ensure that society's recorded memory can be created in an accurate and reliable form, and be authentically maintained and preserved, for short- and long-term use by its creators and by society at large. The first phase of the project examined records created in databases and document management systems, whose preservation was mandated for the accountability and administrative requirements of organisations. InterPARES 2, which began in 2002 and will be completed in 2006, has focused on digital records produced in dynamic, interactive and experiential systems in the course of artistic, scientific and electronic government activities. Dynamic records are those whose content is dependent on continuously varying data, held in several databases and spreadsheets, while interactive records are created and maintained in interactive systems in which each user's entry instigates a response from the system, or triggers a particular action. Experiential records are defined by an essence that goes beyond the bits that constitute the digital object to incorporate the behaviour of the rendering system, or at least the interaction between the object and the system. This second phase of the project has necessitated a reassessment of some concepts in the context of systems in which fluidity is intrinsic to the nature of the records and essential to the accomplishment of their purposes as support for the actions of their creators. The issues revolving around the preservation of such digital objects are more complex than have been

---

[57] DigiCULT's first *Thematic Issue* deals with issues of integrity and authenticity in detail. The publication is free to download from DigiCULT's Web site at http://www.digicult.info.

[58] Researchers have backgrounds in: Archival Science, Chemistry, Computer Engineering, Computer Science, Dance, Diplomatics, Film, Geography, History, Information Studies, Law, Library Science, Linguistics, Media Studies, Music, Performance Art, Photography and Theatre. The countries actively involved are: Canada, United States, Australia, Belgium, China, France, Ireland, Italy, Japan, Netherlands, Portugal, Singapore, Spain and the United Kingdom. The Advisory Board also includes an archivist from South Africa.

encountered with traditional electronic records because the authenticity of their rendering is dependent not only on the objects themselves, but on the reproduction of their relationships to the environments in which the objects are experienced and acted upon.

Although based on interdisciplinary methods and approaches, the project is intended to be archival in purpose; thus, the results of the research activities carried out across disciplines are constantly translated into archival terms, associated with archival concepts, and developed into archival methods. In other words, the ultimate goal of the project is to articulate the characteristics, components and requirements of a trusted system for the creation and maintenance of digital records, and a preservation system that ensures the authenticity of their authentic copies over the long term. Nonetheless, upon completion of the research, the outcomes will be translated back into the language of each discipline so that these systems will be made comprehensible and accessible to records creators in the fields involved with the project. InterPARES researchers have retained the traditional archival definition of a record as any document created (i.e. made or received and set aside for action or reference) by a physical or juridical person in the course of a practical activity as an instrument and by-product of such activity. However, the characteristics established as essential for identifying the records and situating them in context have had to be reconsidered in the light of the nature of the digital objects under investigation.[59] For instance, fixity of form and stability of content are not characteristics of experiential, interactive or dynamic records, and the challenge has been to determine methods for preserving these records either in their dynamic state or in some alternative fixed form and stable content developed by the creator for the purposes of its activities.

CASE STUDIES AND GENERAL STUDIES

Research has been developed around a multi-method design, based on the methodology and tools considered most appropriate by each investigating team for the specific research activity it carries out. A wide range of case studies from the three focus areas (arts, science and e-government) forms the basis of the investigation of current practice with regard to the creation and maintenance of experiential, interactive or dynamic records. Examples include: *Obsessed Again…* an interactive electronic musical work written by Canadian composer Keith Hamel; the geographical information system developed and used by the Centre for Desert Archaeology in Tucson, Arizona; the Alsace-Moselle land registry system; the Antarctic Treaty Searchable Database; and the work of Stelarc, a multimedia performance artist who integrates digital media in his works performed in diverse environments. Each of the three focus areas has presented specific challenges based on the ways in which records are viewed within the particular context, the types of actions that the records are supporting or in which they participate, and the emphasis placed on representation of the records in their original form. For instance, in the arts, both visual and performing, authenticity is conceptualised as the ability of the digital object to reflect the artist's original intentions with regard to both content and form upon every subsequent presentation of the work. In the realm of science, creators are more typically concerned with ensuring the accuracy of data entered into, and maintained in, the systems than with the form of their aggregations, although the authenticity of scientific records is crucial to ensuring the integrity of the data within them. In e-government, accountability is of greater concern than in the other two focuses, but this often leads to confusion over the difference between authentication (a declaration of authenticity at a single moment in time) and the authenticity of the records through time. To date, eleven of the twenty-one case studies have been completed and represented in activity and entity models, and half of these have been analysed according to diplomatic principles in order to gain an understanding of the boundaries of the records contained in each system and to determine their essential characteristics.

To complement the case studies, general studies have been undertaken to address the issues relevant to each of the three areas of records creation. Examples of these include surveys of government Web sites, of the record-making and recordkeeping practices of digital photographers, composers and film makers, of the practice of preservation of interactive music, and of file formats and encoding languages used for non-textual materials, as well as a study of the San Diego Supercomputer Center's project to develop a prototype for persistent archives based on data grids for the National Archives and Records Administration (NARA). At the same time, three teams of researchers responsible for investigating the key concepts of accuracy, reliability and authenticity as understood within the disciplines encompassed by each of the focuses have produced annotated bibliographies and literature reviews, undertaken analysis of these concepts as discussed in the literature, and created bibliographic databases for the management of references.

TERMINOLOGY

A terminology research group has been working towards the standardisation of the vocabulary within InterPARES, so that all researchers and research assistants will be consistent in their usage of terms, thereby ensuring effectiveness of communication within the project, and consistency in dissemination activities. The work of the team has focused on the creation of lexicographic instruments: a Register of all terms and phrases used within InterPARES; a Dictionary including the definitions provided for those terms by all the disciplines involved in the project; a Glossary including the definition for each term chosen by InterPARES as the authoritative meaning for all the project's documents; and a Thesaurus that assigns the terms to a specific facet (e.g. agent, action, object, event, property), links it to other facets through hierarchical, equivalence or associative relationships, and points to the preferred term. The terminology group has worked closely with all other research units in order to ensure the greatest inclusivity while maintaining consistency and rigour of communication, and to keep the delicate balance between accurately describing the nature of concepts used across the various disciplines and prescribing a common language for the research project as a whole. The group is particularly aware of the political nature of its endeavours and of the partiality and bias potentially involved in the prescription of terms and the standardisation of meaning – a big challenge if one considers the encompassing nature of the research, both in terms of the diversity of cultures involved, and of the multiplicity of meanings often associated with terms used in the vast range of fields incorporated within the project. At present, the terminology research is focused on the evaluation of its lexicographic instruments, in order to assess their compliance with international standards, and their ability to meet the needs of external users. Future goals centre on the promotion of public access to the terminology database in order that researchers will be more aware of impediments to effective communication and of possibilities for increased

---

[59] InterPARES 1 identified the following characteristics of a record: (1) a fixed form, meaning that the entity's binary content must be stored so that it remains complete and unaltered, and its message can be rendered with the same documentary form it had when first set aside; (2) an unchangeable content; (3) explicit linkages to other records within or outside the digital system, through a classification code or other unique identifier; (4) an identifiable administrative context; (5) an author, addressee and writer; and (6) an action, in which the record participates or which the record supports either procedurally or as part of the decision-making process.

consistency and understanding among disparate fields of study.

MODELLING

The project has incorporated the use of modelling as a means to conceptually represent the activities and entities under investigation in order to assist in the analysis of case study data and to communicate findings related to creation, maintenance and preservation of the records under examination. The major efforts of the modelling team have been channelled into the development of a model which integrates the appraisal and preservation activity models from InterPARES 1 with the activity model produced by a previous research endeavour, commonly known as the UBC-DoD research project.[60] The amalgamated model (Manage the Chain of Preservation, or MCP model) depicts all of the activities involved in the management of electronic records throughout their lifecycle, from creation to permanent preservation. The model presents a clear visual representation of the ideal management process from the point of view of the preserver, providing the conceptual framework in which the realities of current practice may be measured. The model assists all other units of research by providing a means to identify preservation issues related to the case studies (which are walked through the model and validate it), and to assess the points of the records' lifecycle that require development of policies, procedures and standards (including metadata schemas and reproduction requirements). The group is currently engaged in testing the model and has begun to develop an activity model of preservation from the viewpoint of the creator.

POLICY

The policy research team has endeavoured to produce a conceptual framework for the development of policies that will impact each of the three focuses of investigation. The major area of enquiry centres on the identification of barriers to preservation that currently exist in laws, regulations, policies and standards on copyright and intellectual rights, privacy and freedom of information, authenticity and authentication, open standards and Open Source, and records and archival management. Legislation and standards originating in Canada, the United States, Australia, Singapore, Hong Kong and the European Union (as well as specific

---

[60] Luciana Duranti, Terry Eastwood and Heather MacNeil, *Preservation of the Integrity of Electronic Records* (Dordrecht: Kluwer Academic Publishing, 2002).

countries within it), which have a bearing on records creation, maintenance and preservation, are being analysed within the conceptual framework in order to formulate model policies and standards. Case studies have aided in the identification of areas in which policies are absent or insufficient to ensure the preservation of authentic electronic records created in the arts, science and e-government. Preliminary findings have indicated the need to return to the articulation of the concept of record within each environment – an endeavour that is currently a major focus of discussion among the researchers. A series of model policies and guidelines will then be able to ensure that the entities identified as records are managed appropriately throughout their lifecycles. One objective of this team is to establish liaisons with policy and standards-issuing organisations in order to ensure their effective implementation.

DESCRIPTION

The main objective of the research team focusing on the description of digital entities has been to evaluate existing and emerging metadata schemas, descriptive standards and metadata tools to determine their ability to meet requirements relating to the creation, management, appraisal, preservation and use of reliable and authentic records in the three areas. The purpose of this endeavour has been to recommend either the development of new schemas, standards and tools, or the extension of existing and emerging ones.

The resultant activities of this research group have consisted of a number of specific projects. The team has begun the analysis of the completed case studies in order to identify the various uses of metadata and other descriptive standards within the diverse spheres of activity for the purposes of identifying existing practices, setting a framework for the delineation of best practice, and determining those areas in which current standards are unable to meet the requirements identified. A specific product which is intended to both facilitate this research and be a direct deliverable of the analysis is a Metadata Schema Registry. The registry is a centralised repository of schemas, which will aid various professions and organisations to identify metadata sets, or the combinations of elements from several sets that are appropriate to serve their recordkeeping needs. As it currently exists, the database has nearly completed the design phase of its development, with implementation taking place in the summer of 2005. At present, approximately ten of the most relevant metadata schemas have been analysed with reference to ISO recordkeeping standards, InterPARES 1 requirements for the presumption of the authenticity of records, and the Australian

Recordkeeping Metadata Schema, while a number of other schemas have been identified for registration and analysis.

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

CONCLUSION

The large quantity of research material already generated by InterPARES 2 has effectively laid the groundwork for the last two years of the project, which will focus on the development of its deliverables – namely, guidelines for records creators and preservers, prototypes of appraisal and preservation systems, frameworks for the development of policies, strategies and legislation, descriptive schemas for digital entities, an existing metadata schema registry, and literature and terminology databases.

As mentioned previously, the experiential, interactive and dynamic nature of the digital entities under investigation has necessitated a re-examination of the conceptual foundations of archival theory, the meaning of authenticity in such a technological environment, and what it means to preserve digital records by reproducing them in an authentic way. To return to the analogy of the performing arts: in that context, preservation has traditionally been accomplished by keeping the script or the score, which, to varying levels of detail, specifies the creator's intention of how a work is to be reproduced. Due to the fragility of digital media and the rapid rate of obsolescence of digital technology, the current hypothesis is that the electronic records investigated in InterPARES 2 will require increasingly detailed scripts in order to ensure their authentic reproduction. Identity metadata inextricably bound to the data in the record, linked to a model of its form and a description of its digital environment (i.e. record functionality and system documentation), stored in a trusted recordkeeping system, will potentially be the only means for preserving the authenticity of records, thereby ensuring their continuing reliability and accuracy. While explorations of these processes and systems are still under way, to a large

extent the project has already been tremendously successful, as it has promoted the collaboration of scholars and professionals from a range of cultural and disciplinary backgrounds, devoted to the preservation of their digital culture.

TRENDS IN DIGITAL REPOSITORY MANAGEMENT

The PREMIS Working Group (Preservation Metadata: Implementation Strategies, http://www.oclc.org/research/projects /pmwg/) released a report in September 2004 entitled *Implementing Preservation Repositories For Digital Materials: Current Practice And Emerging Trends In The Cultural Heritage Community*. The report presents the conclusions of an international survey on existing and future repositories for digital materials. The survey focused on current practices in managing metadata in digital archives, but also covered such key aspects as preservation strategy, funding, mission, and access policy. Discussion at the end of the report identifies emerging trends within the sector.

PREMIS is sponsored by OCLC (http://www.oclc.org/) and RLG (http://www.rlg.org/). The full report can be downloaded from http://www.oclc.org/research/projects /pmwg/surveyreport.pdf.

JISC DIGITAL REPOSITORIES PROGRAMME

The Joint Information Systems Committee (JISC) (http://www.jisc.ac.uk/) has announced a programme aimed at facilitating the set-up and use of digital repositories by the UK's education and research communities. The programme will consider how repositories interact at an institutional, national and international level; the range of educational content held in digital repositories (for example, ePrints, learning objects, technical reports, multimedia and datasets); and the informal use of repositories. It will focus on: projects which explore the differing roles of digital repositories (especially where the repository brings together different groups of people); new technologies and tools; pilot services, for example supporting information discovery, or services that could be shared in many different repositories; the development of standards, specifications and frameworks; supporting studies, such as evaluation, or related subject areas, e.g. IPR or data integrity and authenticity.

Related work and programmes by JISC include: FAIR (Focus on Access to Institutional Resources, http://www.jisc.ac.uk/index.cfm?name=programme_fair); X4L (Exchange for Learning, http://www.jisc.ac.uk/index.cfm?name=programme_x4l); MLEs for Lifelong Learning (http://www.jisc.ac.uk/index.cfm?name=programme_mle_lifelong2); Digital Libraries and the Classroom (http://www.jisc.ac.uk/index.cfm?name=programme_dlitc); Linking Digital Libraries with VLEs (http://www.jisc.ac.uk/index.cfm?name=programme_divle); and Digital Preservation and Asset Management in Institutions (http://www.jisc.ac.uk/index.cfm?name=funding_circular4_04).

Funding available through the programme is expected to total around £2.5M (€3.615M) per year, which will be available to proposals from UK HE institutions and their partners.

OPEN ACCESS TO RESEARCH

During 2004, a report *Scientific Publications: Free for All?*[1] was published in the UK to address the fact that scientific research is typically published through specialist journals, to many of which individuals and even academic libraries cannot afford to subscribe. This obvious obstacle to the dissemination of research findings both within and beyond the scientific community is a growing concern in the sector, and many have investigated open access as a solution.[1]

PubMed Central (http://www.pubmedcentral.nih.gov/) is a digital archive of life sciences and biomedical journal materials that is free to access online, funded by the USA's National Institutes of Health. At present, the National Library of Medicine (http://www.nlm.nih.gov/) is working with the Wellcome Trust (http://www.wellcome.ac.uk/) to create a European version of this resource. It is foreseen that depositing research articles in the online archive will become a requirement for those benefiting from Wellcome Trust research grants, and that additional funding will be supplied in order to cover the cost of producing PubMed-compliant metadata and the charges of open access publishers such as the Public Library of Science (http://www.publiclibraryofscience.org/) and BioMed Central (http://www.biomedcentral.com/).

More information is available from http://www.wellcome.ac.uk/doc_WTX022826.html and http://www.wellcome.ac.uk/doc_WTX022412.html.