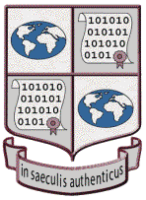
PRESERVING AUTHENTIC ELECTRONIC RECORDS: PRELIMINARY RESEARCH FINDINGS

Proceedings from an International Symposium

February 17, 2001
University of British Columbia

edited by Luigi Sarno

Vancouver, Canada, August 2001



This volume presents the proceedings of the International Symposium "How Do You Know It's the Real Thing 2? Preserving Authentic Electronic Records", which took place on February 17, 2001 at the University of British Columbia (UBC) in Vancouver, Canada.

The Symposium was co-organized by the InterPARES Project and the Istituto Italiano di Cultura in Vancouver, with the support of the UBC Institute for European Studies and of the Italian Chamber of Commerce of B.C.

TABLE OF CONTENTS

Introductory Remarks

Margherita Repetto Alaia*

Director, Istituto Italiano di Cultura, Vancouver

The goal of the InterPARES project is to "develop the theoretical and methodological knowledge essential to the permanent preservation of records generated electronically, and, on the basis of this knowledge, to formulate model international, national, and organizational policies and strategies, as well as standards, capable of ensuring their preservation."

Research in this area was motivated by the consequences of the adoption of new information technologies on the part of governments, corporations and individuals. New dangers, such as technological obsolescence, systems' incompatibility, storage media fragility, and the difficulty of maintaining hybrid systems of record-keeping, were threatening both the existence and the authenticity of the records that must be preserved permanently for administrative, legal, and cultural

---

* Margherita Repetto Alaia is the Director of the Istituto Italiano di Cultura in Vancouver, one of many cultural offices founded abroad by the Italian Ministry of Foreign Affairs for the promotion of Italian culture and civilization. Prior to this, Repetto Alaia directed the Istituto Italiano di Cultura in Washington, D.C. (1994-1996) and taught at Columbia University from 1983 to 1993. A historian by trade,

reasons. This situation risked compromising the accountability of governments and businesses, undermining the productivity of industries and the reliability of commercial transactions, and reducing the confidence of scholars in electronic records as reliable and authentic sources for research. Straddling the private and public sectors, as well as being both international in scope and interdisciplinary in nature, the problem required innovative, collaborative, and multicultural research efforts. Thus, Luciana Duranti, Chair of the Master of Archival Studies at UBC, with the financial support of the Social Sciences and Humanities Research Council of Canada, and of the UBC Hampton Fund, Vice-President Research Fund, and Dean of Arts Fund, assembled a group of researchers spanning four continents and organized in national and multinational teams: the Canadian Team, the American Team, the Australian Team, the Asian Team, the European Team, the Italian Team and the Global Industry Team, all of them comprising members of government institutions, private organizations and the industry, and involving several disciplines. The reason why I listed the Italian Team as a separate entity from the European Team is not to emphasize the Italian participation in the research, but to reflect the actual division of the Teams, which is based on its significant and separate financial contribution to the research. In fact, the Italian Team hosts one International Team Workshop per year in Rome, with the support of the Ministry for Cultural Properties and Activities of the Italian

---

she has authored articles and essays in her field of specialization, that is modern European and Italian history.

Government. In addition, its members, who are in significant numbers, are supported directly by Italian institutions such as the National Research Council, the National Archival Association, and the University of Urbino, and indirectly by the Ministries of Finances and Justice. Because of this context, the Italian Cultural Institute decided to promote the InterPARES Project by contributing to the dissemination of the research and of its findings through the organization of this symposium both last year and this year. The primary purpose of the symposium was to bring to the community at large this specialized body of research on issues that affect all sectors of society and also each of us as individuals. This is accomplished not only by inviting the public to this presentation, but also by publishing the proceedings. Last year's symposium proceedings were published first in English only; subsequently, the Italian Government published them in a two-language version (Italian and English) for distribution to the Italian community in Vancouver and for distribution in Europe.

Bringing this research project to the various sectors of the community aims at generating a ripple effect and at involving, in contributing to finding the solution to these grave problems, both the communities that have already encountered them, such as the business community, and the communities that have not been very much aware of them, such as the cultural community in general, and the creative one in particular. Just like an engineering firm needs to know for as long as a bridge stands what the elements of its structure are, a composer of digital music needs to be able to retrieve, access and play his

compositions years after he first created them. Both types of records, even if carried forward across technologies, must be proven authentic in order to be useful.

The symposium of last year was meant to describe the project, its research questions, its methodology and the problems that each participating country was facing with respect to the control and preservation of electronic records. One year later, the InterPARES Project is ready to present its preliminary findings and will do so according to the structural layout of the research.

In fact, in order to manage the complexity of the issues affecting the permanent preservation of authentic electronic records and to attain the Project's goal, the InterPARES Project was divided into four interrelated domains of investigation. Each domain, supported by a dedicated Task Force of international researchers from a variety of disciplinary backgrounds, addresses a distinct research objective and a dedicated set of research questions. The Domains are Authenticity, Appraisal, Preservation and Strategy. To provide support to all research domains, a Committee composed of members of each Task Force ensures the all the terminology used within each domain is consistent and rigorous by maintaining a glossary.

Each Task Force and the Glossary Committee will report on the work done to date, its methodology and deliverables. The symposium will be concluded by the project director, Luciana Duranti, who will

outline the next phases of the research, and by Ian Wilson, the National Archivist of Canada and incoming President of the International Council on Archives, who will present his view of the impact of the project on the archival community worldwide.

The InterPARES Project and the Italian Cultural Institute in Vancouver wish to thank the Institute for European Studies and the Italian Chamber of Commerce of British Columbia for their help in promoting the symposium, and all the speakers for accepting to extend their stay in Vancouver in order to participate to this event.

Conceptualizing an Authentic Electronic Record: the Development of a Template for Analysis

Heather MacNeil
University of British Columbia, Canada

Heather MacNeil teaches archival studies in the School of Library, Archival and Information Studies at the University of British Columbia. She chairs the Authenticity Task Force within the InterPARES Project.

The objective of the Authenticity Task Force is to determine the conceptual requirements for the preservation of authentic electronic records. The first step in achieving that objective is to identify and define the elements of an electronic record that are relevant to a consideration of its authenticity. The authenticity of a record is assessed in relation to its identity and its integrity. Proving the authenticity of an electronic record implies the need to preserve its identity and integrity over time and across technologies.

The disciplinary perspectives that have shaped the identification of the elements are diplomatics and archival science. Viewed from these perspectives, an electronic record, like its traditional counterpart, is a complex of elements and their relationships. It possesses a number of identifiable characteristics, among them: a fixed documentary form, a

stable content, an archival bond with other records either inside or outside the system, and an identifiable context. It participates in or supports an action, either procedurally or as part of the decision-making process, and at least three persons (author, writer, and addressee) are involved in its creation.

In a traditional record-keeping environment, these characteristics manifest themselves in both explicit and implicit ways. For example, the archival bond may be expressed in a classification code or some other unique identifier that appears on the face of a record. The names of the author and addressee typically appear in the "to" and "from" fields in a memorandum. The name of the author may appear in the letterhead in other types of records. The action or matter to which the record relates is typically expressed in a subject line in a textual record or in a caption in a visual record.

The working hypothesis of the Task Force is that, while they may manifest themselves in different ways, these same or similar elements are present in electronic records either explicitly or implicitly. To test that hypothesis, the Task Force has created a *Template for Analysis*, which is a decomposition of an electronic record into its constituent elements, and which defines each element, explains its purpose and indicates whether, and to what extent, that element is instrumental in verifying the record's authenticity.

Before discussing the individual elements, I should make clear that the *Template* is a generalized representation of an electronic record developed for the purpose of identifying all its known elements. It is not expected that any single electronic record will, or should, include all the elements. The absence or presence of one or more of them in a specific instance will depend on the record's purpose.

The elements of an electronic record included in the *Template for Analysis* fall into four main categories:

- documentary form
- annotations
- context
- medium.

The elements examined under **documentary form** and **annotations** are those that are (conceptually at least) part of the record, i.e., they are visible on the face of the record, or embedded in it, or closely linked to it.

The elements examined in the category of **context** are those that are related to, but outside of, the record, i.e., they are part of the larger documentary and administrative framework in which the records are created, maintained, and used.

Medium is considered to reside both inside and outside the record.

Documentary form possesses both intrinsic and extrinsic elements. Intrinsic elements are the discursive elements within the record that communicate the action in which it participates and its immediate context. These elements fall into three groups: i) elements that convey aspects of the record's juridical and administrative context (e.g., the name of the author, addressee, the date,); ii) elements that communicate the action itself (e.g., the indication and description of the action or matter); and iii) elements that convey aspects of the record's documentary context and its means of validation (e.g., the name of the writer, the attestation, the corroboration).

Extrinsic elements refer to specific, perceivable features of the record that are instrumental in communicating and achieving the purpose for which it was created. For electronic records these include overall presentation features (i.e., textual, graphic, image, sound, some combination of these), specific presentation features (special layouts, hyperlinks, colours, sample rate of sound files, etc.) electronic signatures, electronic seals (e.g., digital signatures), digital time stamps, and other special signs (e.g., digital watermarks, organization crest, a personal logo).

Annotations, which are additions made to a record after it has been created, constitute the next category of elements included in the *Template for Analysis*. They fall into three basic groups. The first group includes additions made to the record after its creation as part of its execution, e.g., the date and time of transmission added to an e-

mail record at the moment it is sent, or the indication of attachments added before it is transmitted. The second group includes additions made to the record in the course of handling the business matter in which the record participates, e.g., comments noted on the face of the record, or embedded in it, and dates of transmission to other offices. The third group includes additions made to the record in the course of handling it for records management purposes. Such additions typically include the classification code or file number assigned to the record, its draft and/or version number, cross-references to other records, an indication of scheduling actions, and so on.

The manner in which annotations manifest themselves in an electronic record-keeping environment depends on the application being used. For example, word-processing applications typically provide for the insertion of comments into a record, along with the identification of the individual making the comment and the date. In records management applications, annotations such as the classification code or scheduling actions are included in the profile associated with the record.

The examination of a record's **context** shifts the analysis away from the record itself to the broader legal, administrative, and procedural frameworks in which the record is created and managed. Examining these frameworks helps us to understand, among other things, the business processes in the course of which electronic records are created, maintained, and used, the types of records generated from

these processes, and the connection between those processes and the creator's broader functions and mandate. That understanding in turn provides a foundation on which to identify more precisely the kinds of supporting documentation and information that are essential to support the verification of a record's authenticity over time and which must therefore be preserved and transferred along with the records when they become inactive and are transferred to the record preserver.

The final category of elements in the *Template* is **medium**. It is taken for granted that a record is a representation of a fact that is memorialized on a physical carrier, i.e., a medium, and preserved by a physical or juridical person in the course of carrying out its activities[1]. From this assumption it follows that a record cannot exist before its elements have been inscribed on or affixed to a medium. Similarly, in an electronic environment, the bitstream, i.e., the source of the record, cannot endure for any length of time unless it is affixed to a medium.

However, as the Preservation Task Force Chair has continually reminded us, with electronic records, storage of a bitstream on a hard, floppy, or optical disk, or on a magnetic tape, may be necessary for the bitstream to endure but it is not sufficient to represent the content and form of a record. Representation of an electronic record's content

[1] Maria Guercio, "Principi, metodi e strumenti per la formazione, conservazione e utilizzo dei documenti archivistici in ambiente digitale," *Archivi per la storia* XII, 1-2 (1999): 26.

and form also requires the capacity to process the record through software[2]. Moreover, while affixing a bitstream to a medium is considered an essential pre-condition to the existence of an electronic record, this does not mean that the medium is an essential or even a relevant factor in verifying that record's authenticity. It is assumed that it is neutral with respect to the record's authenticity at least from the perspective of the records creator and the records preserver.

The validity of the *Template* as an instrument for identifying the elements of an electronic record in general and those specific elements that will enable us to verify its authenticity is being tested through four rounds of case studies of electronic systems that either create, or have the potential to create, electronic records. The remaining presentations in this session focus on the methods and findings of the first three rounds of case studies.

---

[2] Ken Thibodeau, "Certifying Authenticity of Electronic Records: Interim Report of the Chair of the Preservation Task Force to the InterPARES International Team," unpublished report, 19 April 2000, 1.

Understanding the Boundaries of the Electronic Record: Interim Report on the InterPARES Case Studies

Anne Gilliland-Swetland
University of California at Los Angeles, United States

Anne Gilliland-Swetland is an Assistant Professor in the Department of Information Studies of the Graduate School of Education and Information Studies at UCLA. Her teaching and research interests relate to electronics records administration and the development and evaluation of information systems containing primary sources. Anne is a member of the Authenticity Task Force of the InterPARES Project.

Abstract

This paper reports on the electronic records case studies currently being conducted by the InterPARES Project. The paper discusses why the case study approach was adopted; reviews the nature of the data being collected and how it is being analysed; and reflects upon how this data can be used to understand the scope of electronic records within a diversity of national, juridical, institutional, procedural, and documentary contexts.

Introduction

Contemporary archival theory and practice relies heavily on many under-explored assumptions about the nature of records. This lack of

reflexivity permeates the creation, management, preservation, and use of records, where pragmatic considerations often result in records being treated more as non-problematic physical "objects" than as complex, highly contingent intellectual and socio-cultural constructs that they also represent. This lack of problematization of the construct of the record presents critical issues for electronic records researchers who need to be able to understand the boundaries that delimit the record not only in terms of the record as a physical unit of analysis, but also in terms of its contingencies and functionalities, as well as the values attributed to it by different user communities.

The *Template for Analysis* developed by the InterPARES Project's Authenticity Task Force and delineated in Heather MacNeil's paper develops a model of an ideal record that identifies all the possible *known* elements that a record may contain and how those elements might be manifested in electronic environments. This template builds upon prior diplomatic practice where typologies have been developed by analysing what was known about recognized types of records and record-keeping practices, many of which have been in use for hundreds of years. The goal in InterPARES, however, is to develop a predictive model that will assist archivists and others responsible for preserving authentic electronic records (those in existence today, and those that might be developed in the future) by identifying a hierarchy of requirements necessary for maintaining the identity and integrity of different record types over time. A key intellectual consideration in doing this is how InterPARES researchers can validate and, if

necessary, extend upon the *Template for Analysis* in order to account for record types that are yet to be developed. In other words, how can we extrapolate from known characteristics of records to those that might be unknown given that the electronic records and record-keeping systems in use today represent at best imperfect examples?

This paper reports on the progress of the case study method that we have employed in order to provide data to investigate the diplomatic hypotheses around which the project is based. The paper reviews the nature of the data being collected and how it is being analysed; and reflects upon how this data can be used to understand the scope of electronic records within a diversity of national, juridical, institutional, procedural, and documentary contexts.

The Use of Case Studies

The decision to employ a case study method was not something that we had anticipated at the outset of the project. Given the growing complexity of electronic records due to technological developments such as distributed processing and hybrid systems, however, we were faced with a need to understand more definitively the scope of electronic records and the various ways in which those records might be manifested. To address this need, we have adopted a grounded theory approach where theory developed from what is known about traditional records is iteratively re-examined and refined based upon analysis of purposefully selected case studies of electronic record-

keeping systems. In order to ensure consistency in how the case studies are conducted, we have developed a set of case study data gathering and data analysis instruments which includes the *Case Study Interview Protocol* (CSIP) and the *Template Element Data Gathering Instrument* (TEDGI). The CSIP is broken down into five components, based upon each of the contexts already outlined. A range of predominantly open-ended questions are asked to elucidate each context, the elements expected in a record (that is, those contained in the TEDGI), and ultimately to populate the *Template for Analysis*. The same question may also be asked in different ways within the same contextual section to check for consistency in responses. The same question may be asked in a different way in more than one section to try to elicit different perspectives from the respondents. The interview data and other documentation gathered during the case studies allow us to see the interplay between those contexts. Even if questions are answered by one interviewee, they will be asked again of another in order to check for consistency and for different perspectives.

The criteria for case study selection and the instrumentation are modified and refined between rounds in order to i) target data that will best test the evolving *Template for Analysis* and ii) develop a toolset that can be used by other archivists engaged in electronic records appraisal and preservation. For example, the criteria for the second round of case studies were the following:

1. Systems whose contexts indicate a strong likelihood that they contain, generate, or have the potential or possibility of generating records.
2. Systems that have gone through one or more migrations.
3. Systems where migrations were from one electronic system to another electronic system.
4. Systems for which several aspects of technological context (storage media, system software, application software, data format, schema) was changed in the course of each migration.
5. Systems for which both the pre-migration and the post-migration versions are available and are up and running.
6. Systems for which detailed documentation (design, implementation, migration, metadata) exists.
7. Systems with a diversity of information configurations, especially those that least resemble traditional record-keeping systems and/or records (e.g., geographic information systems, Web applications, and large-scale databases).
8. Systems within the same institution that exhibit a diversity in content and type of records.
9. Record-keeping systems that perform similar functions but that are located within different institutional contexts (for example, university student records systems).

In the first round of case studies we also included some inactive systems. We realized, however, that we could learn little of the technological and procedural contexts in which we were interested

from a system that was no longer active, and we decided that all successive rounds would only focus on active systems.

By conducting rounds of case studies, we are addressing several methodological and intellectual concerns:

- *How can we build a predictive model from observed, imperfect implementations of electronic record-keeping and with no way to have knowledge of technologies and processes yet to be developed?*

  The case studies allow us to identify the degree of consistency and variability within the juridical-administrative, procedural, documentary, and technological contexts of records. This gives us an indication of where and how changes are occurring between traditional and electronic record-keeping. The case studies also enable us to develop a flexible model based upon characteristics of electronic records and systems that can be populated by existing and emerging systems.

- *How can we identify conceptually which of a diversity of extant record elements are necessary to demonstrate and preserve authenticity so that future archivists, system designers, record-keepers, courts, and historical scholars can identify them regardless of the range of ways in which they might be manifested?*

By looking across a range of case studies, we are able to observe patterns that may occur in all case studies, or only in those that represent similar record-keeping or technological implementations. The case studies help us to identify the various ways in which different elements might be manifested and, equally importantly perhaps, to see when expected elements are missing.

- **How can we identify what constitutes a record and a record type?**

  The case study method allows us to see the ways in which records exist at many levels of granularity within and across systems. We have been looking at the parameters and status of the record, for example, is it a datafile, an aggregation of records within a system, or the system itself? In some cases, a system may contain one or more types of records, as well as a diversity of non-record materials. In the case of databases, electronic records frequently comprise selected data elements that are pulled together by activity-related metadata such as audit trails, reports, and views through a process prescribed by the business function for a purpose that is juridically required. As part of this analysis, we have been asking at what point does a record become complete, and is it ever fixed and stable? The case studies also elucidate the relationships between electronic and paper components of the

same record in terms of their differing degrees of proximity to the actions in which they participate.

Four successive rounds of case studies are being conducted on systems containing electronic records or that have the potential to generate electronic records. Three rounds, comprising twenty-five extensive case studies in government, industry, academic, and small business settings in Canada, the United States, the United Kingdom, Italy, and China have been conducted and completely or partially analysed to date. A final round of case studies is currently underway which will provide additional data to walk through the models being developed by the other InterPARES Task Forces, as well as to provide cases around which to base educational products such as professional courses and workshops.

Each case study yields several types of data: transcripts of tape-recorded interviews with record-keepers and systems personnel; interviewees' responses to CSIP questions; and supporting contextual documentation such as mission statements, organization and workflow charts, institutional policies, codebooks, and system schematics. Three main sites — the University of California, Los Angeles (UCLA), the University of British Columbia (UBC), and the University at Albany, State University of New York (SUNY-Albany) are compiling, organizing, and encoding the case study data to prepare it for analysis. Diplomatic analysis of case study data is being conducted at UBC, and a range of social science analytical methods

are being applied at UCLA and SUNY-Albany. By conducting different types of analyses, we are able to view the data from a range of vantage points to see if there are other things we can see in it (including understanding what data we need that we have failed to identify with our tools thus far). In this respect, we are not only analysing the data, but also the method, since we have been evaluating how the toolset needs to be refined as well as delineating what kinds of data collection archivists need to do in order to be able to make informed appraisal and preservation decisions. This is a key component for translating outcomes of InterPARES research into practice.

UCLA is testing additional models developed by other electronic records research initiatives against the data in order to understand whether we can adopt, adapt, augment, or reject work that has already been conducted, but from very different perspectives. For example, business process, business function, and systems analysis modeling might help us to identify interdependencies between the record, the system, and the business process.

Another type of analysis we are conducting is narrative analysis of transcripts of interviews, especially the components of the interviews that employed open-ended questions and asked for commentary. In doing this, we are looking for narrative themes and rhetorical tropes that convey beliefs, perspectives, and vocabulary of the different

stakeholders such as records creators and information technology personnel responsible for designing and maintaining the systems. We are also looking at how they are using and understanding words such as "records," "system," and "authenticity", and how they understand what we are asking during the interview.

UCLA is also comparing across similar cases such as student records systems and registry systems in a range of administrative settings in order to understand which contexts may cause variability where the record's function remains reasonably consistent. One way of doing this is by looking for recurring clusters of elements and processes that might provide cues about the types and characteristics of records being examined. If we understand the roles played by such clusters, then we may be able to derive from this understanding a minimal set of metadata elements that can support the continued authenticity of records in electronic systems.

Summary

The case study approach rigorously assesses the project's underlying theory to see the extent to which it translates directly into the electronic environment, and the extent to which it needs augmentation in order to address unanticipated phenomena such as new manifestations of elements of form or different technological and procedural processes. Archivists tend to talk about electronic records in relation to systems – this is inevitable, perhaps, since ultimately

they must manage and demonstrate the authenticity of the records as physical as well as intellectual entities. However, the case study method also problematizes the nature of the system and how it is technologically, juridically, procedurally, and documentarily delimited. Looking at the range of contexts in which the records and systems are embedded helps us to understand our unit of analysis and identify where there are variables within these contexts that might be significant in terms of identifying and implementing authenticity requirements. This activity also supports the InterPARES goal of developing a framework for standards, best practices and policies for different environments, sectors, and situations.

_____

An Overview of InterPARES Case Studies done by the National Archives of Canada

Yvette Hackett
National Archives of Canada

Yvette Hackett is the Electronic Records' Officer in the Government Archives Branch of the National Archives of Canada. The position was created in1994 to provide a central focus for electronic records' issues in the division and to assist archivists in the appraisal, processing and description of electronic records of the federal government. Prior to this, Yvette worked as an archivist, database administrator and descriptive standards officer in the Moving Image and Sound Archives Division at the National Archives. Yvette is a member of the Authenticity Task Force of the InterPARES Project.

The National Archives of Canada has completed six case studies for InterPARES over the last eighteen months. It is a measure of our commitment to the project, and our interest in its research agenda. We realized quite early that, as a national institution mandated to preserve both private and public-sector records, we were in a privileged position to access the wide variety of automated systems needed to support the work of the Authenticity Task Force. We also had a second advantage: a group of experienced and knowledgeable archivists who have worked with electronic records through most their careers -- specifically Catherine Bailey, David Brown, Jana Buhlmann, Marnie Burnham, Cara Downey, Andrew Horrall and Paul

Marsden.

Our case studies have focussed on government systems that we had already appraised as having long-term value. In brief, the six case studies investigated were:

- a geographic information system
- a registration system (database)
- two records management systems

and two quite new systems:

- one containing personnel records
- and one containing case files.

One of my colleagues, Catherine Bailey, on analysing our selections, discerned a pattern in our choices, which could be roughly classified into three types (but I hasten to add, these are not diplomatic types). The first is the traditional or "classic" computer system - typically a large database running on a mainframe or minicomputer. The second type can be referred to as "middle" or developing systems - those which combine advances in technology with elements of traditional record-keeping. Finally, the third type includes the "cutting edge" systems - those which are not only technically advanced but have begun to illustrate record-keeping practices in a potentially paperless environment.

I will briefly describe the salient features which attracted us to each of these systems, and identify the findings which we felt would be most

useful to the InterPARES researchers in formulating authenticity requirements. Our first two case studies qualify as "classics". The first was one of the earliest geographic information systems in the world, called the Canada Land Data System, which had already been transferred to the National Archives. The application was already dormant when our preservation efforts began, and it presented the National Archives with significant technical challenges related to software obsolescence and non-standard data formats. Unfortunately, the fact that the system was no longer operational impeded our ability to fully represent its role in its original operational environment, which is a key element of an InterPARES case study.

Our second "classic" case study focussed on the Indian Registration System, operated by the Department of Indian Affairs and Northern Development. Under Section 5 of the *Indian Act*, the Department is required to maintain an Indian Register, which records the name of every person who meets the definition of "Indian" as set out in the Act. Since 1985, the Indian Registration System (IRS) is that Register. As relational database applications go, it is a fairly straightforward design and is typical of hundreds of similar systems across the Canadian federal government. Where it differs from the majority is in its strong and clearly defined legislative context. The IRS provided early evidence of the most common security protocols that protect databases, such as physical security, password protection, and audit trails. But the application itself was lacking in both extrinsic and intrinsic elements of form - no logos, no watermarks, no digital

signatures, and no overt identification of authors, addressees or archival bonds.

For the final case study in Round 1, we studied the electronic records management system operating at the National Parole Board (NPB). The Parole Board is an independent administrative tribunal that makes conditional release and pardon decisions, and recommendations for the exercise of the Royal Prerogative of Mercy. The Board exercises exclusive authority over parole and certain aspects of conditional release of federal offenders, and over parole of provincial offenders in the ten Canadian provinces and territories that do not maintain their own parole boards. Its automated records management system supports all business functions of the Board through the control of its paper and electronic records.

There were several reasons that the NA offered this "middle" system as a case study candidate. First, electronic records management systems are becoming increasingly common within the Canadian federal government, and it was felt that this type of system should be addressed in the earliest stages of the case study research. The National Archives had appraised this system in 1994 and issued its first Terms and Conditions for Transfer written for an electronic records management system. In addition, the NA was familiar with the FOREMOST software itself, which complies with the U.S. Department of Defence's Records Management Application Design Criteria Standard 5015.2, which itself reflects the findings of the first

University of British Columbia research project on the Preservation of the Integrity of Electronic Records.

NPB's records management system offered several elements of traditional records management, such as a subject-based block numeric classification system and textual record forms such as memos, correspondence and reports, while also offering several new twists, such as a document profile which consistently records information about each document, automated version control, and a linkage system for attachments.

One important issue relating to electronic records management systems is that they offer the implementing organization many configuration choices that could affect the resulting store of documents. To further investigate the impact of these "implementation choices," we studied a second records management system running in the Pacific Region offices of the Department of Indian Affairs and Northern Development. This system integrates both document management and records management software, and includes a scanning component for incoming correspondence.

The last two case studies I will discuss involved "case file" systems that have long presented problems of volume and granularity to archives attempting to capture the relationship between government and its citizens. Early database applications associated with case file

series were largely tracking systems with limited tombstone information coupled with a "bring forward" function. Today, the newest systems aim to replace or reduce the paper case file itself, with a combination of scanning and the direct capture of records in electronic form.

The Personnel Electronic Records Management Information System (PERMIS) went live at the Department of National Defence on April 1, 2000. Until that time, DND had maintained a system of paper-based "career" files for each individual serving in the Canadian Armed Forces. In the past, duplicate files were frequently generated as personnel were deployed across Canada and around the world. The new electronic files contained in PERMIS replace all active paper case files. They are constructed almost entirely of scanned documents, with some personal identification data imported from the payroll system. PERMIS offers a common scenario for scanning-based systems, where the original paper forms are records that generate actions (such as transfers and promotions) and can be annotated. The paper records are then scanned at a central location, at which point the image of the paper form becomes the "official record" and the paper is destroyed.

Our final case study was the Canadian Automated Patents System (known as TechSource), built by the Canadian Intellectual Property Office (CIPO). The system supports all the processes required to issue a Canadian Patent. In 1997, the National Archives, while

acknowledging the archival value of the Canadian case files, had refused to accept electronic transfers from TechSource, preferring the more complete and stable paper case files. In 1999, the system had evolved and CIPO requested a re-evaluation. With the technical evaluation already in progress, we expanded the research and completed a case study for InterPARES.

The system description that follows is accurate to the period in late 1999 when it was collected, rather than to the current system which continues to evolve. The various documents comprising a TechSource case file use most standard methods of electronic records creation. Patent application forms are received either in hard copy or in electronic form. Paper applications are scanned, while electronic applications are converted to a long-term logical format. Control information concerning each application is keyed into database tables and linked to the scanned images, and subject-based indexing is performed. Additional database tables assign tasks to employees and record updates to the status of the application.

Incoming correspondence is also largely paper-based, but all outgoing correspondence is generated electronically. Form letters are highly automated and are created by merging information from relevant database tables with “standard clauses” which are subject to version control. Finally, customized correspondence is generated in a word-processing software and the final result converted back into TechSource.

TechSource was designed with long-term retention as a primary criterion, since a patent file remains active for 20 years. It also must be capable of receiving input and generating output which meets the World Intellectual Property Organization's defined exchange formats. From the perspective of the InterPARES case study, the system's most interesting features involved the tension between a record being "set aside" while remaining in a live application, and the relationship between the various individual documents and database occurrences and their "aggregate" existence as a file.

In conclusion, I would like to take this opportunity to publicly thank the Canadian federal government departments who agreed to participate in these case studies. By allowing strangers to scrutinize their systems, they have contributed a great deal to the InterPARES research.

_____

The Growing Complexity of Music Preservation

Brent Lee
Post-doctoral fellow, University of British Columbia, Canada

Brent Lee is currently a Post-doctoral fellow within the School of Music at the University of British Columbia, and a researcher on the InterPARES Project. He holds degrees from McGill University (B. Mus. 1986, M. Mus. 1990) and from the University of British Columbia (D.M.A., 1999). In addition to his research into the nature and challenges of preserving digital music documents, he is an active composer of orchestral, chamber, and electro-acoustic music, and a developer of music educational software.

The theme of my presentation today is the growing complexity of music preservation. It seems paradoxical that in a era of easy digital duplication that the continued existence of musical works, or more specifically, the documents that enable us to hear musical works in the same manner that they were created, should be somehow in question. Yet, the results of case studies and interviews conducted within the sphere of InterPARES research reveal the preservation of much of the music made today to be a perplexing challenge for individual musicians, archival institutions, and the cultural industry at large.

I began a presentation made to a group of new media artists last year by asking if they had, in the past, created performance-based works

that could probably not be recreated today. Every one of the artists assembled held up their hand. Several of my own works composed in the late 1980's using various combinations of synthesizers and software are similarly unperformable: a tragic loss to humanity. Another composer interviewed for an InterPARES case study reported enormous obstacles in remounting works written as little as five years ago -- a work written ten years ago would not be possible to perform again. Why is this the case? What has happened that has made the preservation of some music so fragile?

The digital revolution has affected the processes by which music is created to the same extent as it has affected other processes in other disciplines. Today, most of the music that we hear has been recorded or remastered digitally, notated digitally, and/or performed wholly or in part on digital instruments. At the avant-garde of music composition, composers are developing interactive environments where computers accompany live performers in an intelligent way, improvising responses to the actions of a performer within parameters predetermined by the composer. The traditional model of a composer writing down music for interpreters to perform on acoustic instruments is still a common one, but for many musicians the recording studio and the desktop computer have opened new avenues of artistic exploration and, in turn, placed in question the strategy of preserving music by preserving the scores and sketches of composers.

Of course, the advent of sound recording more than a century ago

represented an entirely new way in which music might be preserved. Originally, recordings documented musical performances; it was only with the invention of magnetic tape that the editing of recordings became commonplace, and a new art form, the studio recording, was born. In the late 1940's composers began experimenting with this new art form, constructing collages of recorded sound; this compositional model did not require a performer in the traditional sense, and is today commonly termed electroacoustic music.[1]

Thus at the end of the pre-digital era of manuscripts and analogue recordings (if such an era can be said to have ended), music could be preserved on paper (scores, sketches, and prose descriptions of music) and on tape or vinyl (performances, studio recordings, and electroacoustic compositions).[2] The vast majority of the current holdings of musical archives are just so: paper records and analogue recordings,[3] as is the pre-1980 back catalogue of commercial record labels.

In the 1950's, composers began to use computers to synthesize compositions and even to "compose" music according to algorithms

[1] In practice, electroacoustic works may be "diffused" in live performances, that is, an interpreter (usually the composer) controls the volume and spatial location of the sound as the tape is replayed.

[2] Of course many other media were used for analogue recording, including wax cylinders and discs made of acetate, vulcanite or shellac.

[3] See, for example, the archival holdings of the Music Division of the National Library of Canada: http://www.nlc-bnc.ca/services/emusicar.htm (address current as of 2001/02/07).

designed by the composer. The digitally synthesized compositions usually fell into the broad category of "electronic music" as they were ultimately recorded on tape, yet algorithms were used not only to control the digital synthesis in some way but also to create scores for acoustic instruments. Interestingly, the archival funds of Lejaren Hiller, an American composer who used the ILLIAC computer at the University of Illinois to generate the compositional material for a number of his works, contains only textual records, photographs and films, and analogue recordings. The algorithms he used to compose his music are only broadly represented on flow charts.[4] Of course, whether this is a tragic loss to humanity or a tragic loss only to the legion of musicologists devoting their careers to the study of algorithmic composition is a matter for debate.

Perhaps one is tempted be a bit flippant about the preservation of Lejaren Hiller's algorithms because for Hiller the goal was to create musical scores, which he did, and to have those scores performed and recorded, which they were. These scores and recordings are well preserved, and certainly represent the core of his work. But when documents at the core of a musician's work are digital in nature, preservation issues become more complex.

As recording technology shifted from analogue to digital in the

[4] The archival description of the Hiller *fonds*, preserved at the Music Library of the State University of New York - Buffalo is available at: http://ublib.buffalo.edu/libraries/units/music/spcoll/lhhome.html (current as of

1980's, musicians and record labels expected the same standard of preservation afforded by analogue media. Unfortunately, it became apparent that this standard was a very low one; the shelf life of analogue tapes is limited to a few decades or even less if not stored under ideal conditions. Still, with analogue tapes, deterioration is incremental; a number of strategies for coping with dimensional changes and the chemical decomposition of the binding agent have been developed.
Thus, an analogue recording can in most cases be restored, at least to a certain extent.

Conversely, digital recordings tend to be either playable or unplayable. The most common formats for digital recording, DAT (digital audio tape) and U-matic tape, have both shown the same types of disintegration as tape used for analogue recordings, but the problems are compounded due to the thinness of the digital tape, its plastic housing, and the inability of the error correction mechanisms built into the recorders to cope with damaged tape.

To be sure, the problems associated with the instability of storage media are common to the preservation of all electronic records. For some musicians, the instability of the entire environment in which digital records are created compounds the problem of preservation.

A number of new technologies were developed in the 1980's that

---

2001/02/07).

allowed for more complex interaction between computers and people. Of particular interest to composers was the possibility of the computer processing a variety of inputs, including sound (through microphones) and movement (through motion sensors). Composers began to write "interactive" pieces in which the computer "improvised" a musical accompaniment to a performance by a musician or dancer. In other words, the algorithm-based composition pioneered by Lejaren Hiller became something that happened during a performance, and also became an essential part of the performance: the fact that no two performances were identical or even similar appealed to many composers and performers alike. Such performances employ a computer running algorithmic software and in turn controlling a rack of synthesizers and sound processors; the authentic preservation of this music depends not only on the preservation of the digital records generated in the process of producing the original performance, but also the algorithmic software, operating system, computer peripherals, and various musical equipment.

And here is the difficulty of which our music case studies have given evidence: each of these diverse system components has a life span, and in most cases, a short one. It may be possible to migrate the records from one multi-purpose computer to another, but each synthesizer and sound processor has its own operating system and its own proprietary set of file formats. Musical equipment manufacturers rely on the obsolescence of older models to drive the sales of newer

models; the newer models are rarely backward compatible. Further, even if a historic piece of equipment could be reverse-engineered, its design would remain the intellectual property of the original equipment manufacturer, and its duplication a patent violation. Thus the fragility of works composed in this medium: if I wished to remount a performance of my brilliant 1984 opus "Mr. Crumb" I could probably find an Apple computer with a MIDI interface, I may be able to find the Session 8 software I used, but the chances of finding the necessary synthesizers in working order are close to zero.

As you struggle to come to terms with this further tragic loss to humanity, I ask you to consider the potential parallels between algorithmic composition such as it is and records creation such as it might be in coming years. An obvious parallel exists between the functionality inherent in algorithmic composition records and the functionality inherent in web pages. The embedding of *Java* script and applets allows web pages to behave differently depending on the input of the user. As the terms and conditions surrounding an Internet transaction are articulated in the web page, or are "just a few clicks away", the web page may have to be preserved, with all of its functionality, so that an accurate record of the transaction can be created.

A second parallel lies in the reliance of web pages on propriety applications in the form of plug-ins: the continued intelligibility of the website (and thus records generated by interaction with the website)

depends upon the continued functionality of these third-party applications. Just as with synthesizers and sound processors, such applications evolve rapidly and rely on obsolescence to boost future sales. As more and more of our interaction with businesses and government agencies is via computer, and as the interfaces between people and computers become even more sophisticated, the preservation of the entire environment surrounding a transaction will become necessary to reproduce the record of that transaction.

I would like to close with a brief comment on one aspect of music preservation that has become apparent to me in the course of conducting InterPARES research, that is, the inability of individual musicians and small institutions to manage their digital records. The arrangement and description of personal archives has always been potentially problematic, but these problems may soon be overshadowed by quantities of unreadable or unintelligible digital materials: notation files, digital recordings, algorithms, and synthesizer patches. Individual musicians often lack the skills, resources, and inclination to carefully manage their records -- any time spent doing so is time not spent making music. I imagine that this situation is paralleled in the personal archives of writers, artists, and any individual that uses a computer. As many of these records are not relied upon for accountability or even for reference, the urgency of their preservation is qualified. Still, the long-term cultural and historical value of the records of our artists in all disciplines seems to me to justify a great deal of concern.

We expect that the current InterPARES research will resolve a number of these issues, and that some of the more vexatious issues raised by the growing complexity of music-related documents might be dealt with in the next round of the research. While "Mr. Crumb" may be lost forever - tragically lost forever -, I hope that pieces composed today might stand a better chance of permanent preservation.

Draft Conceptual Requirements for Authenticity

Luciana Duranti
Director, InterPARES Project

Luciana Duranti is a Professor in the Master of Archival Studies Program at the School of Library, Archival and Information Studies of the University of British Columbia, where she has taught since 1987 and where she occupied the position of Associate Dean Research for the Faculty of Arts from 1997 to 1999. Prior to moving to Canada to teach at UBC, Luciana Duranti - a former President of the Society of American Archivists - taught at the University in Rome and served as State Archivist in the State Archives of Italy. Professor Duranti holds a Doctorate in Arts (1973) and graduate degrees in Archival Science from the University of Rome (1975) and in Archivistics, Paleography and Diplomatics from the School of Archivistics, Paleography and Diplomatics of the State Archives of Rome (1979).
Duranti, who received the Canadian Universities Faculty Association of British Columbia's 1999 Academic of the Year Award, is presently the Director of the InterPARES Project, as well as a member of the Authenticity Task Force.

On the basis of the analysis of the case studies and of the fundamental assumption of the Preservation Task Force that it is not possible to preserve electronic records, only to preserve the ability to render them, the Authenticity Task Force has drafted baseline requirements that should allow for the preservation and verification of the authenticity of all types of electronic records.

When referring to records, authenticity refers to the fact that a record is what it purports to be and has not been tampered with or otherwise corrupted. Authenticity is therefore the trustworthiness of the record as a record. It is to be distinguished from reliability, which is the trustworthiness of a record as a statement of fact, as to its content that is, and from authentication, which is a declaration of authenticity, resulting either by the insertion or the addition of an element or a statement to a record, and a means of proving that a record is what it purports to be at a given moment in time.

To verify that a record is authentic, one must verify its identity and integrity. The identity of a record is provided by its provenance, author, addressee, writer, date, matter or action, and archival bond (i.e., relationship to the records that belong in the same aggregations). The integrity of a record is its wholeness and soundness. A record has integrity if it is intact and uncorrupted. This state is not absolute: records on every medium show the sign of the passage of time. Therefore, an electronic record is to be considered intact and uncorrupted if the elements conveying its identity are intelligible and the message it is meant to communicate in order to achieve its purpose is unaltered (i.e. if the articulation of the content and the formal elements required for the record in question are the same as when the record was first set aside).

The authenticity of records in live systems is threatened during transmission across space or time, and during their maintenance by the creator, especially when this involves migration from an

obsolescent to a new technology. Therefore, it is essential that the records be handled and maintained in the context of a trust management system, that is, of a record-keeping system such that it allows establishing a presumption of authenticity for the records under its control. The records of the creator that would then be considered authentic are those that exist as created, as they have not undergone processing, and any copies of them that result from a migration process.

The need for a trust management system is predicated on the basis of the argument that inference of trustworthiness for the records of the creator derives from the fact that the creator treats its records as trustworthy by relying on them for action or reference in the usual and ordinary course of business. However, such inference is not supportable when the records are no longer actively used because the motivation to maintain them accurate and authentic is diminished. Hence, it is essential that the designated preserver, before accepting the records of the creator in its custody, be able to verify that they are clearly identifiable and of demonstrable integrity, and that accidental corruption or purposeful tampering have not occurred after the records are no longer in active use by the creator.

On the basis of this reasoning, the Authenticity Task Force has established baseline requirements for records creators, which will allow the record preserver to attest that the records were maintained in a trusted record-keeping system and to ascertain their identity.

According to such requirements, the preserver must verify that the creator has:

1. Defined access privileges.

2. Defined rules by which electronic records move inside and outside the organization by determining:
   a) what are the data that flow into the record profile for incoming, outgoing and internal records in accordance with their state of transmission (whether they are originals, drafts or copies);
   b) the routing of the records received by the organization and sent by the organization.

3. Established the routine according to which the electronic system will generate a record profile for each record. The profile should serve the purpose of an annotation and be linked inextricably to the record.

4. Designed a profile including fields that allow the verification of:
   a) identity of the record:
      i) Persons (author, writer, addressee, etc.)
      ii) Action or Matter
      iii) Dates (of the document and archival)
      iv) Expression of archival bond (classification code, dossier identifier, etc.)
   b) integrity of the record:

i) Name of Handling Office
ii) Additions
iii) Deletions
iv) Migrations (date of migration)

5. Established methods and rules for authentication (a means for verifying the authenticity of the record in time) by:
   a) linking to the integrated business and documentary procedures the authentication requirements for specific types of records and assigning responsibility to officers or offices for implementing the requirements;
   b) assigning to the records office the competence for issuing authentic copies of electronic records;
   c) establishing the authentication procedure for each form of conversion of records;
   d) creating a view of the records profiles of all the records within a dossier or class.

6. Established rules for copying by:
   a) identifying the need for generating records in multiple copies on the exclusive basis of working needs and vital records needs;
   b) identifying the various types of copies (e.g., simple transcription, imitative copy, copy in the form of original, authentic copy, insert);

c) identifying the authority to be accorded to each type of copy;

d) establishing procedures for routine copying of records that are needed beyond the life expectancy of their medium (in the case of electronic records preserved outside the system).

7. Established a tracking and location system by:
   a) when a closed dossier is removed from the system, assigning a location which is recorded within the electronic system according to class;
   b) establishing a procedure for maintaining up-to-date location information for all active and semi-active records outside the live system.

8. Established audit procedures by:
   a) maintaining an audit trail of access to the records system to control the administration and use of access privileges;
   b) maintaining an audit trail of every transmission (date, time, persons, action or matter) within the record system.

9. Established procedures to prevent loss or corruption of records because of intentional or inadvertent unauthorized additions, deletions, or alterations by:
   a) providing the electronic system with the capability to restrict access to the back-up procedures to authorized personnel;
   b) prescribing that back-up copies of records and their profiles be made periodically;

c) maintaining an audit trail of additions and changes to records since the last periodic back-up. It should contain the information necessary to provide recovery of records in the event of system failure and it should be maintained on an electronic system different from the one containing the records;

d) maintaining a system back-up that includes system programs, operating system files, etc.;

e) ensuring that, following any system failure, the back-up and recovery procedures will automatically guarantee that all complete updates (records and any control information such as indexes required to access the records) contained in the audit are reflected in the rebuilt files and will also guarantee that any incomplete operation is backed up. The capability should be provided to rebuild forward from any back-up copy, using the back-up copy and all subsequent audit trails.

10. Established procedures to prevent the loss of records due to factors such as technological obsolescence (of hardware, system software, and storage media such as: storage devices, access methods, and database management system) by:
    a) planning upgrades to the organization's technology base;
    b) ensuring the ability to retrieve, access and use stored records when components of the electronic system are changed;
    c) migrating records.

11. Established procedure for taking records out of the live system for preservation purposes by:
    a) identifying the officers authorized to remove records from the system;
    b) determining storage medium and location for records removed from the system;
    c) determining what has to be removed along with the records (e.g., indexes, data directories, data dictionaries, profiles, etc.);
    d) using the retention schedule to implement the transition of records from active to semi-active status and from semi-active to inactive status;
    e) determining methods of transfer of inactive records to the entity competent for their preservation and the form in which the records will be transferred.

12. Established records storage facilities and equipment requirements by:
    a) forecasting the rate of accumulation of active and semi-active records taken out of the live system;
    b) determining space and climate control requirements;
    c) determining the need for records storage facilities and equipment.

13. Compiled all the above policies and ensured that every migration procedure that has occurred has been properly documented outside of the system.

If the record-keeping system of the creator satisfies these baseline requirements, the preserver can accept the creator's records into its custody and assume responsibility for an ongoing reproduction that guarantees the continuing authenticity of the records. Authentic preservation of electronic records is thus ensured by the ongoing production by a neutral third party, the designated preserver, of authentic copies of authentic records.[1] A copy of a record is considered authentic if it is declared to be authentic by a person officially entrusted with such function. This declaration of authenticity, this authentication that is, is both constitutive and probative of authenticity as it both makes a reproduction authentic and proves its authenticity.

The verification and attestation of the authenticity of the copies of electronic records by the preserver is predicated again on the general notion of trust management in that the preserver takes up the role of trusted custodian and must meet foundation requirements that support

[1] The term copy is here used in a general sense of reproduction. A copy is thus any reproduction of a record. It can be made from an original or from a copy of either an original or another copy. There are several types of copy. The most reliable copy is the copy in form of original, which is identical to the original, although generated subsequently. An imitative copy is a copy that reproduces both the content and form of the record, but in such a way that it is always possible to tell the copy from the original. A simple copy is a copy that only reproduces the content of the original. An insert is a simple copy included in a new record. A record may have two other states of transmission: it can be a draft or an original. A draft of a record is a temporary compilation made for purposes of correction, while an original is the first, complete record which is capable of achieving its purposes (i.e., it is effective). An authentic copy is any of the types of copy listed, which is declared to conform to the original by an officer entrusted with such a function.

the presumption of authenticity of the copies declared authentic. The baseline requirements for the preserver are:

a) the preserver shall fully document the activity of reproduction, demonstrating both the relationship between the records acquired from the creator and those reproduced, and the impact of the technology chosen for the preserved copies on the form, content, accessibility and use of the records;

b) the date of each reproduction and the name of the responsible person shall be introduced in the proper field in the record profile of each record;

c) the identity of the record must be clearly expressed by preserving where appropriate (e.g., in the record profile, on the face of the record, in a register) the elements that are necessary to determine it: at a minimum, the names of the persons involved in the creation of the record, the action or matter, the date of the record, and the expression of the archival bond;

d) the documentary context must be accessible and clearly understandable both through the preservation of the records aggregates necessary to understand the records in question and through an archival description of the *fonds* at the various levels of records aggregations;

e) the administrative context of the record (juridical, provenancial and procedural) must be clearly understandable both through its description as provided by the preserver and through its documentary context;

f) unbroken custody of the record must be maintained and demonstrated by the preserver;

g) security and control procedures must be implemented and monitored.

These baseline requirements for the preserver establish grounds for a presumption of authenticity of the records in its custody: until proof to the contrary is shown, records that meet the requirements are considered authentic. However, it has to be emphasized again that the preserver's declaration of authenticity is only as strong as the evidence on which such declaration rests. If the identity and integrity of the records have been lost or compromised while they were in the hands of their creator, the preserver can only declare them as authentic as they were at the time of transfer to its custody. This is the reason why it is so important that records be maintained by the creator under the control of a trusted record-keeping system.

The baseline requirements for the creator and the preserver are meant to apply to all kinds of electronic records. However, the Authenticity

Task Force has hypothesized that electronic records can be distinguished by types and that specific requirements for authenticity can be determined for each of those types. Thus the Authenticity Task Force is now in the process of building a typology of electronic records. The methodology for doing so will be illustrated by Seamus Ross.

Building a Typology for Electronic Records*

Seamus Ross
University of Glasgow, U.K.

Seamus Ross is Director of Humanities Computing and Information Management at the University of Glasgow. He runs the Humanities Advanced Technology and Information Institute (HATII), which provides an academic undergraduate and postgraduate programme and carries out research in application of advanced technologies to arts, humanities and heritage sectors.
Seamus is a member of the Authenticity Task Force in the InterPARES Project.

I would like to start my speech today with an example that bears a relation, although not an apparent one, with the focus of my speech, i.e. building a typology for electronic records.

In 1956, Sir Thomas Beecham went to a studio in London and conducted a performance of Tchaikovsky's *1812 Overture*. The recording of that is still in the public domain and you can regularly purchase it to listen to it. About 3 weeks later, he went to Edinburgh, where he conducted the very same piece of music to a live audience, in that year's edition of the Edinburgh Festival. That recording only

* This paper has been transcribed from the audio recording of the Symposium's proceedings.

survives by accident. A lot of effort has gone into preserving it, and a digital version was also created. When you play the two recordings one after the other, a substantial difference appears clear: in the studio recording, Beecham and the orchestra rushed through the performance, while in the live recording they dragged out every moment to bring the emotion and the feeling to the audience itself.

This points to the fact that if we want to understand performance - especially the relationship between the audience and the performers - the preservation of live performance as well as studio performance is absolutely essential. What does this have to do with typology? First of all, they were the same piece of music but the performance was different, since they were two different *types* of performance. If we can classify performances under different types, we can ask different questions about what happens in different kinds of environments and we can classify the questions that we are going to ask based upon the environment in which that activity is taking place. In that instance, it was the difference between a live performance and a studio performance -- a difference that we are now able to understand.

A similar sort of thing is going on in the InterPARES project. We believe that if we can type records, we may be able to learn a lot about how to handle individual categories of records. I am going to talk a little bit about the typological process and our reasoning about it, with reference to our goal, i.e. finding ways to handle records and authenticity in a coherent manner. We are focussed on creating a

typology of electronic records related to authenticity, with special attention to identity and integrity. The ability to group records into particular types should provide a tool to improve our ability to handle electronic records and their preservation. But first let me explain about the typologies and the process of creating them. This should give you an indication as to why the design of the CSIP and the TEDGI (which Anne Gilliland-Swetland talked about this morning), working practices and methods of analysis that we are using at UBC and UCLA are absolutely central to the work of the Authenticity Task Force.

Typification is a process of abstraction and analysis. It involves the design and testing of the definitions of classes through assigning entities, whether physical or conceptual, to these classes. The goal of typification is the development of a framework for structuring entities, whether objects or even abstract concepts, into homogeneous, correlated groups. These groups need only be related by a degree of family resemblance, but they are most likely to involve a combination of hierarchical and relational classifications. We should produce a framework that can accommodate the typoes of records we know, but also new types of records, and also types of records we don't even know exist. This is because we do not want to build a typology that is based upon a framework of only the known, we want to be able to predict the kinds of new records that are going to come along and be able to slot them into our framework. When you are doing typological work, only very rarely do you actually work with a closed and

bounded set, to which no additions will ever be made. As a consequence of our approach - and of the fact that InterPARES is operating against the backdrop of a rapidly changing technological and business environment -, the final result of our work should be a predictive model that allows for new types of records to be identified and added. It is our belief, and one demonstrated by the creation and use of typological frameworks in other areas, that once we know that a physical object or an entity belongs to a particular type, we will automatically know certain things about that entity. In the case of material culture, say objects, you may know about the temporal and spatial distribution of the objects themselves. In the case of records, if we get our types right, we may know something about the processes and systems that were involved in their creation. In what ways will authenticity requirements manifest themselves in those records? The purpose for which the typology is being developed is the key to the ways in which it will be structured. The aims of the typology we are creating relate closely to the goals not just of the project but of the Authenticity Task Force itself. This is important because there may be more than one way to classify the same sets of objects; also, a typology is not built simply by looking at a set of materials and deciding their attributes and the values to be assigned to those attributes. You have to actually make sure that you are choosing the significant attributes and the meaningful values for them. I am going to give you two examples that hopefully will help understand what we are doing in a less than abstract way.

Imagine for a moment that you are a waiter in a restaurant. I come in the restaurant and ask for a bottle of Perrier, but you don't have any, so you name half a dozen other kinds of mineral water. The only mineral water I'm familiar with is Perrier though, I don't know which of those is going to be like Perrier. If you are an experienced waiter, you will have actually tasted Perrier as well as all the other mineral waters you serve. You therefore know that Perrier has large, heavy, while other waters have smaller but hard bubbles and others still very soft bubbles. Knowing that the mineral water I prefer is Perrier you can probably serve me some soda water as Perrier, with its hard and aggressive bubbles. You did not have to know about the specific object but all you had to have was a way of classifying it. Now, this process happens daily to all of us, making us much more functional and making it much easier for us to handle information.

Before I was involved in InterPARES I had never attempted to build a typology for conceptual objects. I build typologies for material culture and that's a very different set of processes because with material culture, the objects are actually visible, while in this case our job is much harder. The underlying and absolutely essential element of this process has to do with how we are selecting our data and the process by which we are performing the data collection. An inadequate typology is naturally bound to result from inadequate data collection: this is why we have put a tremendous amount of emphasis on the design of the Case Study Interview Protocol (CSIP) and the Template Element Data Gathering Instrument (TEDGI).

Very often, when people talk about typology, they talk about the analysis stage. Like so many other things, the analysis stage has to be founded by extremely high-quality data. So, we are basically looking at records as entities: in other words, entities can be pins, pots or lamps or, as is the case with our work, abstract concepts.

Let me give you another example to help understand the process of typifying imaginary beings. I would like you to think of the Jimmy Stewart movie with Jimmy Stewart's wonderful friend Harvey. Harvey is real to Jimmy Stewart. You never see Harvey but you come to believe Harvey is a real being so Harvey has certain attributes. The first attribute that Harvey has is that you never see him. Or, to take yet another example, think of Calvin and Hobbes. Calvin, as you know, is the little boy who has his stuffed tiger. His stuffed tiger is an imaginary being who is quite real to Calvin, and very real to us as the readers of the comic strip, but just a stuffed tiger to everyone else in the comic strip, like his parents. Now, out of this class, this "type" of imaginary beings, we have on the one hand a sub-type related to Harvey and similar kinds of imaginary individuals that you never see, and on the other hand we have a sub-type of imaginary friends that are like Hobbes who are only real to those of us who are prepared to believe they are. In other words, we have taken a set of entities which are these imaginary creatures, we have chosen a set of attributes about them (e.g. being seen/unseen) and we have assigned values to those attributes. Once we've done this, based upon those values we are able

to group our imaginary friends. This type of process applied to electronic records ought to allow us to handle the material better, to understand the processes that underlay their production. I would like to point out two other key issues that are absolutely essential in conducting typological studies. One is that in choosing the attributes, we have to make sure that they are meaningful to the creator of the record. In the case of the Harvey and Hobbes examples I made, you can see that the attributes I chose were quite meaningful to the creator of the comic strip and to the director of the movie. Still, these attributes also need to be meaningful to the observer. These are two very important premises that underlie our thinking. We have to make sure that the attributes we choose are meaningful to the creator of the record *as well as* to the observer. The observer does not necessarily need to be a person, but could be another computer. Our types will therefore need to have clearly defined and rigid boundaries. One record will belong to only one type. The types will be mutually exclusive, in other words. All the units will be independent and there will be an equal distance between them.

Another important point is that in order to arrive at our classification we use both a top-down and a bottom-up approach. In the examples I gave you before, I combined the two approaches, and I obtained better results. It is clear to us that using only one approach, say the top-down approach, is not going to be effective. Taking a bottom-up approach is not going to be effective either, because we are not going to know what questions to ask. So, the Authenticity Task Force has

been experimenting with typological development which is an iterative process: now that the case studies have provided us with more data to make this modeling possible, much of our time is spent to generate and test scenarios of explanation. Once we are able to type records based on authenticity, we shall know about the processes of their production, and we ought to know at that point what can be done to create authentic copies. The typological work should assist us with sharing the results of the Task Force with the Preservation Task Force, thus creating a seamless stream of understanding that flows from authenticity through appraisal into preservation.

Overview of the Diplomatic Analysis of Electronic Records within the Canadian Automated Patent System (TechSource)

Anna Gibson
University of British Columbia, Canada

Anna Gibson is in her final term of the Master of Archival Studies programme at the School of Library, Archival and Information Studies of the University of British Columbia. Anna completed her Bachelor of Arts degree in Classics at the University of Alberta in Edmonton. She has been a research assistant for the InterPARES Project working on the Authenticity Task Force since September 2000.

The task of performing a diplomatic analysis on electronic records within the systems in the second round of case studies was given to the Master of Archival Studies research assistants at the University of British Columbia. Each case study was assigned to two research assistants, whose first responsibility was to determine whether or not the system contained records. By studying the Case Study Interview Protocol (CSIP), the Template Element Data Gathering Instrument (TEDGI) and additional supporting documentation provided by the researchers with each case study, the research assistants determined which systems contained records. Records were identified based upon the notion that a record must meet the following criteria:

a) a record has at least 3 persons (author, writer, addressee) involved in its creation;
b) a record has a fixed documentary form;
c) a record has an archival bond with other records both inside and outside of the electronic system;
d) a record has stable content;
e) a record has an identifiable context;
f) a record participates in or supports an action either procedurally or as part of the decision making process.

Once it was determined which systems contained electronic records, the research assistants' second responsibility was carrying out a diplomatic analysis of those records. A diplomatic analysis involves identifying the extrinsic and intrinsic elements of documentary form, as well as the procedures involved in the creation of the record. The extrinsic elements of documentary form are those elements that determine a record's material make-up and appearance such as its medium, presentation features, and special signs. Intrinsic elements of documentary form are those elements conveying the action in which the record participates and its context. Through this analysis, the research assistants isolated those elements of the electronic records needing to be maintained in order for the record to remain authentic. From this process, the baseline requirements for authenticity were drafted.

Using this approach, the electronic system in Case Study 19 was identified as one that contains electronic records. The Canadian Automated Patent System, otherwise known as TechSource, contains a variety of records used in the administration and examination of patent applications. Operating under the Canadian Intellectual Property Office (CIPO), TechSource was developed in part to reduce paper control costs and to increase operational efficiency through automation. The system description that follows is accurate to the period in late 1999 when it was collected, rather than to the current system which continues to evolve.

Almost all records that arrive at CIPO in paper format are scanned into TechSource (with exceptions such as oversized documents) and the paper original is destroyed. If records are received in electronic format, they are converted for storage within the system. Two types of records maintained within TechSource are the incoming formal patent application and the written petition for a patent. Both of these records may arrive in electronic format, with the written petition attached to the on-line patent application, or as paper records. The patent application and written petition are maintained as records within the system, and the data from these records are also used to populate tables with information about the patent applicant. These data are used in the course of the patent granting process to generate standard outgoing correspondence from TechSource to the applicant. Standard outgoing correspondence is generated by merging

information about the applicant with the appropriate standard correspondence template and printing it onto letterhead.

Custom outgoing correspondence is also maintained within TechSource. Unlike the standard outgoing correspondence, however, custom outgoing correspondence is word-processed. It is then converted for storage in the system while a paper copy is sent to the intended addressee. Examiner reports are also created as word-processed documents, and converted for storage in TechSource. Finally, most incoming correspondence is still received on paper and is scanned. This process will be reduced, where possible, with procedures that will avoid the "print-to-paper" phase for both CIPO and its clients.

All of these records (i.e. the formal application, the written petition, outgoing correspondence, examiner reports and incoming correspondence) carry a patent application number which links all records pertaining to one patent application. Furthermore, any records that are scanned into the system are automatically associated with the patent number, and the date on which records are scanned into the system is automatically logged. The link between the image and the patent application number is then preserved within the system.

Other components of the system include task lists, subject index links, and audit trails. The task lists assign the various internal actions that are part of the patent granting process to the appropriate work areas.

Subject indexing categorizes all patent applications according to types defined by the World Intellectual Property Organization (WIPO). They are necessary to assist CIPO in checking new patent applications against other patent documents in the system. Audit trails track all of the activity occurring in the databases and any changes, additions or deletions that may appear.

All of the records listed above and the various components of the electronic system participate in the action of administering and protecting patents. Each record within TechSource participates in a different phase of procedure, from the initiative to the execution. The patent case files are dispositive as to function, while audit trails are probative. The author of all records created by TechSource is the Commissioner of Patents, while the addressee and writer may vary. Incoming correspondence may have any number of authors and writers, but the addressee will be the Commissioner of Patents. The creator's assumption of the authenticity of the records within TechSource comes in part from procedural controls established outside of the electronic system. Input into and modification of the information within the system is controlled by the organizational structure of CIPO. Employees are assigned user accounts, which limit the transactions they can perform in TechSource. There are also audit trails and other technological controls for user security.

Currently TechSource meets most of the proposed benchmark requirements for authenticity. The system contains defined access

privileges and established routines according to which the electronic system will generate a link between records. There are established procedures to prevent loss or corruption of records because of intentional or inadvertent unauthorized additions, deletions or alterations. Audit trails of transmissions and of access to the record system are maintained. CIPO has also established procedures for taking records out of the live system for preservation purposes and defined rules by which electronic records move inside and outside of the organization. CIPO has also acquired the necessary record storage facilities and equipment to support their functions.

Other benchmark requirements for authenticity include procedures for the routine copying of records needed beyond the life expectancy of their medium, and procedures to migrate the records to prevent loss over the course of time due to technological obsolescence. These requirements are sufficient to ensure that the records within the system are indeed authentic.

The diplomatic analysis of the records within TechSource was a lengthy but essential process. The collaboration between the research assistants and the case study researchers was invaluable in enhancing the research assistants' understanding of the system. The analysis of all the case studies has contributed to the work of the Authenticity Task Force in formulating the Draft Requirements for Authenticity of electronic records.

Appraising Electronic Records: The Vital First Step in Preserving Authenticity

Terry Eastwood
University of British Columbia, Canada

Terry Eastwood is Associate Professor of Archival Studies at the School of Library, Archival and Information Studies at the University of British Columbia. He is the Chair of the Canadian Team of researchers in InterPARES, and Chair of the Appraisal Task Force. He has written about and teaches the subject of appraisal.

Infused as it is with information technology, our world little thinks of the long-term value of the data, information, and records produced and used to conduct its affairs. It is a world dominated by a seemingly insatiable lust for speed of communication and the latest, supposedly best information, with the result that everything seems transient. Life flies by and is gone. In a strange way, the technology is both cause and effect of the grip that speed and transience have on us. It is cause in the sense that its very existence allows us to create and communicate information faster and faster. It is effect in the sense that it pushes us to get things done now. More and more information passes between us to get those things done, some of it tightly connected with actions, much more simply part of the social ambience. Despite the prodigious capacity of the computer

technology to store information, little room seems to be left for memory of past events and actions. Such technology moves on relentlessly to reform itself, leaving its residue surrounded by lethargy and neglect, or making it the victim of obsolescence. Appraising electronic records, then, is both a matter of determining what we will need and to a certain extent how and whether it can survive to serve our needs.

Archivists began to wrestle with this challenge over twenty years ago, when few others in the world had awoken to devilish ironies. So much information, so tantalizingly available, with so little effort to husband it over the long term. As they witnessed the way organizations created data, information, and records using computer technology, they despaired of ever being able to extract the needful for the long term. The archival literature of the 1980's and 1990's is full of exhortations to organizations to make provision for keeping records over the long term in the processes of systems design and management. Without such, archivists could not effectively dispose of electronic records. Without such, it was not easy to see when and in some cases how things could be extracted from the live systems. Nor was it easy to determine whether the things were indeed records, or whether anything extracted would be the thing it was when it was created, or whether, if extracted, the thing could be preserved in a manner that maintained its trustworthiness.

Organizations intent on preserving electronic records will meet these problems when they face up to the need to evaluate what it is that they want to preserve. Thus, appraisal of electronic records is the vital first step supporting all others in the process of preservation of electronic records. That being so, the Appraisal Task Force of InterPARES aimed to see what could be done to spell out appraisal criteria and methods that would establish a sound foreground in the process of preserving authentic electronic records.

The members of the Task Force are myself as Chair, Barbara Craig of the Faculty of Information Studies at the University of Toronto, Phil Eppard, of the Faculty of Information at The University at Albany of the State University of New York, Gigliola Fioravanti of the Italian Central Direction of Archives, Normand Fortier of the National Archives of Canada, Mark Giguerre of the National Archives and Records Administration of the United States, Ken Hannigan of the National Archives of Ireland, Peter Horsman of the School of Archival Science and Research in the University of Amsterdam, and Du Mei from the central archives administration in China.

We began with a review of the literature and institutional documentation bearing on appraisal of electronic records, the assumption being that the archivists had already encountered most of the main questions we set out to answer. As this assessment of the state of knowledge and experience was under way, we kept a watchful eye on the work of the Authenticity Task Force as it

developed the *Template for Analysis* and the methodology for case studies. We also contributed to the development of the methodology of the case studies and absorbed, from our perspective, the findings of the case studies as they became available. With all this early work in mind, we set to work at the beginning of the second year of the project to produce activity models to get a picture of how appraisal of electronic records takes place. In particular, we were interested in identifying where issues of authenticity are met and addressed in the process. Modeling is not an end in itself but a tool to assist the members of the Task Force to conceive the various stages in the appraisal process for electronic records and what occurs at each stage. The purpose of the modeling, then, is to define the activities involved in appraising authentic electronic records for long-term preservation. We three speakers today would like to summarize the findings this modeling has assisted us to reach so far in our work, which is, of course, not yet finished. I shall begin with some of the more general conclusions we have reached. Normand Fortier and Lara Wilson will discuss our more detailed thinking about the activities we have identified as being most important to the goal of preserving electronic records and their authenticity.

We developed our models from the perspective of the preserver, i.e. the entity responsible for long-term preservation of the authentic electronic records of an organization. The entity may be, to take but a few examples, the archival institution of a government, the archival program or unit of a corporate body like a business firm, or the

archival program of an institution like a university or hospital. We have very consciously tried to generalize the activities to fit any juridical or societal context, any organizational setting, and any administrative setting of the preserver.

From the preserver's perspective, appraisal, i.e. the evaluation of records to determine their disposition, takes place in the context of a broader activity that we call "Select Electronic Records." As is well understood, appraising records to determine those that need to be preserved on a continuing basis involves more than simply making that decision. It means making a selection of records and determining the conditions of their transfer to and acquisition by the preserver. It is also well understood, as reflected in the literature, that electronic records need to be appraised in essentially the same manner as traditional records and in conjunction with any related traditional records. Thus, as with traditional records, appraisers of electronic records will have to take account of societal needs, the record creator's needs, legal requirements, and the principles of archival science. Because the issue of authenticity is heightened in the case of electronic records, we pay special attention to authenticity requirements as developed in archival science and as articulated in the course of the project.

Another way to put it is that evaluation of societal needs, creator's needs, and legal requirements has to be separated from consideration of what needs to be done to ensure authenticity. It is one thing to

decide the continuing value of the records for society, and quite another to assess and apply requirements for authenticity. It is quite possible to want to keep records whose authenticity is doubtful or difficult to preserve. In any case, the determination of both value and authenticity must be made to reach an informed decision about what needs to be done to preserve the records in the long term.

In the broadest sense, selecting electronic records simply means earmarking some for continuing preservation by the preserver, and not selecting others. The process of doing this is mainly one of gathering and evaluating several kinds of information about the records and their context. A major task and a principal expected benefit of our work is to specify the kinds of contextual information, including information about the technological context of records, that needs to be gathered during appraisal. The groundwork to identify the relevant contextual information has in fact been done in the development of the *Template of Analysis* and in the methodology for the conduct of case studies, which have aimed to gather all the information necessary to develop a typology of electronic records. In effect, this work to develop a typology constitutes a methodology for understanding the types of electronic records we face. Our task has been to adapt that methodology to selection.

Thus, we see the selection function as a process of gathering relevant information about the juridical-administrative, provenancial, procedural, documentary, and technological contexts of the records,

and from the electronic records themselves. By juridical-administrative context, we in InterPARES mean the legal and organizational system in which the body creating the records belongs. The provenancial context refers to the creating body itself, its mandate, structure, and functions. The procedural context is the business procedure in the course of which the records are generated. The documentary context refers to the network of relationships of the records with other records of the creator and among the records being examined. Technological context encompasses hardware (including storage medium), software (including operating system, system software, network software and application software), data (including file structure or the relationship of files within a system and data format/file format or the organization of data within files, as well as system models and information about system administration.) This information is then analysed and utilized as the basis for the actions taken at the various stages in the process. The product of gathering, analysing and utilizing this information is very largely documentation about the reasons for making any particular selection and about the records actually selected for preservation. The information issuing from selection as documentation (actually as records of the conduct of selection) is necessary to effect disposition of the electronic records and supports conduct of their preservation. Much of this information about the context of the records needs to be carried forward to the other processes of preservation, because, as is well understood, electronic records do not themselves speak their context in all the

ways needed to understand them once they are removed from that context.

Of course, the preserver must set its selection policies, strategies, procedures, and standards, which then control or guide the activities conducted during the process. Having got a grip on what is involved in these activities, we are now in fact ready to specify the matters that policies, strategies, and procedures for appraising electronic records must address. The two main activities are appraising electronic records and carrying out disposition of them. At this juncture, we may recognize that appraisal of electronic records involves assessing the worth of continuing preservation, on the one hand, and determining the feasibility of preserving authentic electronic records, on the other. Normand Fortier will discuss assessing value and Lara Wilson will discuss determining feasibility.

When electronic records are appraised, then, the relevant information to make a decision has been gathered, analysed, and documented in such a way as to direct carrying out their disposition. Following the preserver's rules and procedures, the responsible agent can then carry out disposition of both records selected and those not selected for preservation. Normally, disposition will be either transfer of custody of the records from the entity creating and/or maintaining them or destroying them.

But now I am beginning to tread on the ground of my fellow panelists. I have set the stage for them. They will be able to give you a more specific idea of where the important analyses of appraisal take place, and the results that each analysis produces.

Appraising Electronic Records

Normand Fortier
National Archives of Canada

Normand Fortier has been an archivist and a Special Projects Officer with the Government Records Branch of the National Archives of Canada since 1992. He has particularly been involved with the appraisal of records created during scientific research. He is also responsible for the Records Disposition Authorities Control System, an instant information system that is used to share disposition resources within the National Archives. Normand is a member of the Appraisal Task Force of the InterPARES Project.

The Appraisal Task Force identified three main activities within its top-level activity, "Select Electronic Records": i) "Manage the Selection Function", ii) "Appraise Electronic Records" and iii) "Carry Out Disposition of Electronic Records". I will describe the second one. The wider context for this presentation is provided in Terry Eastwood's paper.

What is Involved in Appraising?

Appraising a body of electronic records means "to evaluate [these] records for the purposes of continuing preservation" -- in other words, to decide on their disposition. In most cases, this translates into whether they are going to be preserved or not. One of our goals in building the model was to find out where and how authenticity requirements come into play during appraisal.

For the purposes of this presentation, I have assumed that the entity responsible for continuing preservation is distinct from the records creator. That corresponds to the situation of an autonomous archival institution, or, possibly, of an archival unit within a larger records-creating institution. However, the model could most probably be used, in a simplified form, to represent appraisal conducted by the records creator for its own operational purposes. This is something we may look at in the future.

In this Task Force, we have seen appraisal as being made up of four sub-activities; I will now give a quick overview of these activities and how they connect. Please note that the order of presentation does not necessarily imply a temporal sequence of execution.

- *Compile information about electronic records*

  In order to conduct an appraisal, we need information drawn from reading the form and content of the records, and information about the records' various contexts (juridical-administrative, provenancial, procedural, documentary, technological). That information has to be gathered, organized and recorded in order to be used. That is an essential activity: any appraisal must rest on a foundation of solid research.

- *Assess the value of electronic records*

  The second activity we have identified is to assess the value of

electronic records. When we do that, we determine their capacity to serve the continuing interests of their creator and of society. We answer the question: how valuable are these records? How important is it to preserve them? The output of that activity is information about the criteria that were used to assess the value of the records, and how they were applied, as well as an assessment of the authenticity of the records. I will describe this activity in more detail below.

- *Determine the feasibility of preserving electronic records*

  It is not enough to assess the value of electronic records; we must also determine the feasibility of preserving them as authentic records. More precisely, we must decide whether the record elements conferring authenticity can be preserved given our current and anticipated capabilities. In order to determine that, we need the same type of information from the records and about the records that is used to assess their value, but we also need to know our current and anticipated capability to preserve electronic records: this includes the state of preservation knowledge, hardware and software capabilities, staff expertise, and financial resources. That information is actually provided by the preservation function. The output from that activity is information about the cost and technical capability required for continued preservation of the records. This activity will be covered in detail by Lara Wilson.

- *Make the appraisal decision*

  The determination of value and of feasibility come together in the final sub-activity, "Make the appraisal decision". One could see it as balancing what we would like to preserve against what we can afford to. But that is overly simplistic. What we are trying to do is balance value and feasibility. Ultimately, that is a matter of judgment, and it has to be done on a case-by-case basis. Feasibility and value come in degrees. For example, we could be confronted with a situation where preserving records would be either extremely difficult for technical reasons, or would entail prohibitive costs. But this does not necessarily mean we would decide against preserving them. If the records were of extraordinary importance, or if we were required by law to preserve them, we might look for alternate sources of funding, look for another preserver, or come to an arrangement by which the creator would preserve them, at least for certain period. On the other hand, preservation capabilities do come into play, because resources are not infinite and choosing to preserve a given body of records might prevent us from acquiring another set of records later.

The outcome of this decision-making is of course an appraisal decision. But it must also produce a detailed explanation of how and why the decision was reached, both for accountability purposes and so that future users of the records understand better where they came from: out of the universe of records that were created, which ones

were preserved, which ones were not, and why. Finally we must also prepare an information package about the records that were appraised: the information that was compiled about the records and their contexts must be preserved and communicated to those who are responsible for carrying out the disposition of the records, and eventually to those who are responsible for the preservation and the description of the records.

Assessing the Value of Electronic Records

I now come back to the first activity I mentioned, "Assess the value of electronic records". We have decomposed this activity further into three sub-activities: i) Assess the continuing value of electronic records, ii) Assess the authenticity of electronic records, and iii) Determine the value of electronic records.

i) Assess the continuing value of electronic records

The first activity results in a statement of the reasons why the records should or should not be preserved, according to the criteria decided upon by the preserving institution. Because it involves values and judgment, appraisal may be performed differently according to national or intellectual traditions, juridical systems (including legislation), value systems and theoretical choices. Archivists engage in heated debates about appraisal criteria and methodologies, for good reasons. As an example, one could study the records themselves and determine various elements that make up their value: legal value, evidential value, and secondary value,

or their potential value for research or other uses in the future. Another approach, particularly useful when there are vast amounts of records, created during complex, intertwined processes, is to start by appraising not the records themselves but the functions performed by the records creator, to determine which ones should be documented for posterity, and then finding out which records better reflect the accomplishment of these functions and their impact on society. Since our goal here is to come up with a model of the appraisal activity that applies in a number of different contexts, we deliberately omitted specifying which values, strategies or methodologies are going to be employed.

ii) Assess the authenticity of electronic records and

iii) Determine their value

The second activity in assessing the value of electronic records is "Assess the authenticity of the records". We want to establish the trustworthiness of the records: are they what they are supposed to be? Is their identity and their integrity intact? Free from tampering and corruption? This is one point in the appraisal process where authenticity requirements defined by the Authenticity Task Force come into play. The interplay between value and authenticity is not a simple one, however, and there a few important points to note here.

First, authenticity may be ensured in different ways. If I may quote a draft report from the Authenticity Task Force, "identity and

integrity can be verified either on the face of the record, on metadata related to the record, or on components of any of its contexts". It could be through safeguards built into an integrated electronic record-keeping system (the technological context of the records), or it could be insured through policies, procedures and practices in the environment, or around the electronic system, so to speak. Think of physical restrictions on access, policies on access privileges, or procedures for data entry and validation. As a matter of fact at present this is most often what we encounter.

Another important point is that what we are trying to establish is the strength of the presumption of authenticity. Even if all the controls and safeguards are not in place, it does not necessarily mean that the records are corrupted or have been tampered with.

Finally, and perhaps most importantly, the impact of the authenticity of records on their archival value is not straightforward. During appraisal, what we are trying to do is to maintain an accurate reflection of the manner in which the process or function we are trying to document was conducted by the records creator. If the records creator was performing a function that had a very high impact on society, and its record-keeping practices were very poor, we would still want to document the function and acquire relevant records. As a matter of fact, if we knew that the records had been tampered with voluntarily, even fraudulously, this would make it perhaps more important to

preserve the records in order to attest to that.

But then, if the authenticity of the records does not affect their value, why bother assessing it during appraisal? One reason is that it *may* affect the value of the records. Take the case of records that no longer reside in their original environment, that have been migrated into a new one or stored offline. Whether they were originally created and used in an environment that guaranteed their authenticity or not, we still need to know if what we are appraising accurately represents the records as they originally existed. This may not be the case if the chain of custody of the records has been broken, or if a migration has resulted in missing records or parts of records, or inaccurate documentation. If we have good reason to suspect the records no longer reflect what they were at the time of their creation and primary use, we may decide not to preserve them. Another case in which authenticity is important is if the value of the records resides in the accuracy of the information they contain, such as is the case with survey or scientific data, rather than in how well they represent the process during which they were created. Again in this case authenticity is paramount.

Finally, another reason to assess the authenticity of records during appraisal is that this information is crucial to understanding and using the records once they have been transferred to the long-term preserver. In all cases I mentioned, users of the records must know what the presumption of authenticity of the records is -- that is part

of the metadata of the records, so to speak.

Thus, we must assess both the archival value (in the traditional sense of the term) and the authenticity of the records in order to determine their overall value, and decide how important it is to preserve them.

Determining the Feasibility of Preserving Electronic Records

Lara Wilson
University of British Columbia, Canada

Lara Wilson is in her final term of the Master of Archival Studies programme at the School of Library, Archival and Information Studies of the University of British Columbia. She also holds an honours Bachelor of Arts degree in Art History and English, and a Master of Arts in Art History, both from the University of Victoria. Lara Wilson has been a research assistant for the InterPARES Project, working on the Appraisal Task Force, since September 2000.

As Normand discussed, another activity of appraising electronic records involves determining the feasibility of preserving them. This means that the archivist responsible for making the appraisal decision has to decide whether the record elements that confer authenticity can be preserved over time, and in the archival institution's custody. In order to make this decision, the activity of determining the feasibility of preserving **authentic** electronic records is broken down into 3 parts:

1. **first,** the archivist determines the record elements that need to be preserved according to requirements for authenticity; these requirements, already mentioned, were formulated by the InterPARES Authenticity Task Force;

2. **second,** the archivist identifies where these crucial record elements are manifested in electronic record components; and
3. **third,** she or he must then deal with the matter of reconciling these preservation requirements with preservation capabilities of the archives, the institution which is responsible for their continuing preservation. The result of determining the feasibility of preserving authentic electronic records amounts to information about the record components to be preserved and information about preservation feasibility. This information is then used when making the appraisal decision, which in turn affects the disposition of the records. I'll now make further explanation of these activities.

The first activity of determining the feasibility of preserving a given aggregation of records is the identification of intrinsic and extrinsic elements, as well as any data such as audit trails, which need to be preserved in order to maintain the records authenticity over time. Depending on the types of records made or received and set aside by a creator in electronic system, certain elements and accompanying data may or may not be needed to preserve the records' authenticity.

In order to make a determination about preserving these elements, the archivist considers information about the records' various contexts, that is to say the relevant information about electronic records. This is comprised of information about: the legal and organizational system in which the creating body belongs; the creating body's mandate,

structure and functions; the business procedure in the course of which the record is created; the *fonds* to which the records belongs and the *fonds'* internal structure. This structure is the archival bond, the relationship that links each record incrementally to the previous and subsequent ones and which conveys meaning about the records' creation.

There is other relevant information that also aids in determining the feasibility of preserving authentic electronic records. The archivist gathers information from electronic records; this would involve reading the form and content of the records, perhaps by examining samples of the record types considered for acquisition. The preserving agency also needs to know about the technological context of records creation. This would involve gathering and analysing information about the electronic system itself, the hardware, software, operating system and the type of files created – for example, word processing files, image files, etc.

Affecting this determination are appraisal strategies; these are an archival institution's rules and conventions for appraisal. Indeed appraisal strategies are taken into consideration at all three stages in determining the feasibility of preservation.

The activity of determining the record elements to be preserved can be illustrated by using one of the InterPARES case studies. As Anna Gibson and Yvette Hackett explained today, the Canadian Intellectual

Property Office (also known as CIPO) has a system called TechSource, which contains a variety of records produced during the patent-granting process. A very important record is the patent application: it is a dispositive record, that is to say, it is a legal document – the patent application constitutes the first step, or act, in the patent-granting process. The archivist knows this from the consideration of the records' various contexts.

Clearly, in the act of determining the feasibility of preserving authentic records, the archivist would be particularly concerned about this and other dispositive records in the system, and the rigour in which the creator applied authenticity requirements to them. Authenticity requirements set out, for the creator, the measures needed to preserve the identity and integrity of the electronic records. Adherence is the basis for a presumption of authenticity. These requirements are expressed in record elements, whose preservation maintains the records' authenticity over time and across technologies. After considering the records and their contexts, the archivist responsible for appraisal would determine that a dispositive record such as a patent application would have many elements that need to be preserved in order to maintain its authenticity, and therefore trustworthiness. What would result from the archivist's deliberation would be a list of record elements that have to be preserved. In the case of a patent application, certain extrinsic and intrinsic elements such as the application's standard format, and the chronological date and time of receipt by CIPO must be protected from tampering, and,

in fact, are stipulated in the Canadian *Patent Act* (time is important: the time a given application was received by CIPO may have consequences for other applicants vying for similar patents). Intrinsic elements relating to the identity of the record, such as the names of the persons involved and expression of archival bond in the application number, would certainly be included in this list.

Once the elements of the record that need to be preserved have been identified, the next activity is to identify how these elements are manifested electronically as record components. With paper records the extrinsic and intrinsic elements are united on the page, but this is not the case with electronic records. Record components are also known as digital objects - a digital object is (or is a part of) an electronic record that has a method associated with it, such as software. This identification of components is made with information about the record elements to be preserved and with information already gathered about the record's technological context.

For instance, in CIPO's TechSource system, any standard correspondence generated in the course of the patent-granting procedure is produced from templates, with standard phrases and formats, and from table data. These are separate electronic components, which contain the various record elements to be preserved. In order for the archives to be able to preserve electronic records over time, the archivist has to be aware of what these components are, what records elements they contain, and the means

by which the elements can be united to reproduce the record in a comprehensible form.

The final stage of determining the feasibility of preserving authentic electronic records is the reconciliation of the records' preservation requirements with the archives' preservation capabilities. The question is: can the components that manifest these important elements actually be preserved, especially in light of current and anticipated future capabilities? Or, more clearly: realistically, can your institution preserve these records?

This is answered by knowing the archives' current and anticipated capability to preserve electronic records. This information includes the state of preservation knowledge and the institution's hardware/software capabilities, as well as the very practical matters of staff expertise and financial resources available for preservation services.

From this attempt to reconcile preservation requirements with preservation capabilities two bodies of information are produced that inform the appraisal decision. The first concerns the record components to be preserved: this information would explain the way in which the records' elements are manifested in the electronic environment, and would be created to instruct preservation activities. To use the TechSource example again, this information would explain that standard correspondence consists of templates and populated

tables, and would also include their location and how they are identified in the system. The second body of information produced by this reconciliation activity is feasibility information: this would be a statement about the cost to, and technical capability of the archives, required for the continuing preservation of the body of records being considered. Based on current or anticipated finances and technical capabilities, it may be feasible to preserve a given body of authentic electronic records - or perhaps it would not be feasible. Nevertheless, the appraisal decision is made in the light of the institution's appraisal strategies, of the information about the record components to be preserved and about feasibility, as well as of the valuation information articulated in the value assessment activity.

As Normand discussed, the archivist making the appraisal decision produces the appraisal decision, as well as documents containing information about the appraisal decision, and information about the appraised electronic records. These conclusions then inform the disposition of the records.

Preserving Records, Saving Appearances: Perspectives on Digital Preservation

Ken Thibodeau
U.S. National Archives and Records Administration, U.S.A.

Dr. Kenneth Thibodeau is Director of the Electronic Records Archives (ERA) programme at the National Archives and Records Administration (NARA). ERA is a research and development programme aimed at bringing the benefits of advanced computation to preserving and providing long-term access to valuable records. He has 25 years experience in archives and records management, and is an internationally recognized expert in electronics records. Dr. Thibodeau, who earned a Ph.D. in the history and sociology of science from the University of Pennsylvania, served as Chief of the Records Management Branch of the National Institute of Health before coming to NARA in 1988.
Kenneth is a member of the American InterPARES research team and the Chair of the Preservation Task Force of the InterPARES Project.

In an important sense, preserving an electronic record is not a process of keeping the record itself but one of saving its appearance.

In many ways, the process of preserving an electronic record is the opposite of what is done to preserve records on paper or other 'hard' media. Traditionally, assuming the records are on durable physical media, preserving them is virtually equivalent to holding on to what you have. The process of preserving then mostly consists of keeping them in a safe place. One of the biggest risks to the preservation of

records on paper and other traditional media occurs when they are used. Users may damage the records; they may write on them; they may steal them. In extreme cases, records in use may suffer simply from exposure to light or other environmental factors.

With electronic records, the situation is in many ways reversed. Electronic records can readily be protected from risks entailed by use through relatively simple measures, which ensure that users do not get access to preservation media, but only to copies of the records. Standard computer industry practices such as keeping back-ups and using redundant storage systems further ensure the integrity and recoverability of what is stored. With electronic records, however, the assumption of durable physical media is invalid. By and large, information technology has not produced durable digital media. Even if it did, or may in the future, produce durable storage media, the continuing problem of rapid obsolescence would make them irrelevant. These problems of media fragility and continuing obsolescence are well documented. We need not consider them further now; however, regardless of the durability of the media, digital storage systems can be highly reliable. In fact, the need to migrate electronic records across successive generations of storage media can be seen as contributing to their preservation because it entails repeated checks on the integrity of transmission over time. Furthermore, there is clear evidence that in a well-managed system, media migration may not entail additional costs, but be a cost-saving measure. That is due simply to the continued operation of Moore's

law, where storage densities and data transfer rates go up, while unit costs go down. The net result is that migrated records occupy less space and take less time and fewer computing resources to retrieve.

All in all, we can say that, in the case of electronic records, holding on to what you have will make it increasingly difficult, if not impossible, to get it back. Clearly, there would be something wrong with a preservation process that made the records irretrievable. The retrieval of electronic records is a critical concern in preservation. In its *Guide to Managing Electronic Records from an Archival Standpoint*, the Committee on Electronic Records of the International Council on Archives' states:

> The archives must articulate preservation and access requirements to ensure that archival electronic records remain available, accessible, and understandable.

And it goes on to show that preservation and access processes must be tightly linked[1].

The InterPARES Preservation Task Force is charged with identifying and developing the procedures and resources required for implementing the conceptual requirements and criteria for preserving electronic records identified in the first two domains; that is, requirements and criteria for authenticity as applied to the records selected for preservation. The Preservation Task Force is addressing this charge principally by developing a formal model of the process of

[1] International Council on Archives, *Guide for Managing Electronic Records from an*

preserving electronic records. In addition, the Task Force is developing related products, including a dictionary of terms and constructs used in the model; a template for applying the model to specific sets of records; an entity model of the things that are involved in preserving electronic records; and guidelines that institutions can use to articulate comprehensive and coherent frameworks to guide the development and operation of a preservation system specifically tailored to the records each institution is responsible for preserving.

The development of the InterPARES Preservation Task Force process model is a sustained, concerted, international, multi-disciplinary effort to explore in depth everything that is required to transmit authentic electronic records over time. As you will see shortly, we have made enough progress in this effort to support a claim that we will be able to produce an adequate, rigorous, well-founded model that is applicable and adaptable to a wide variety of situations. But before that, we should consider some of the basic concepts underlying this model.

Central to the preservation model is the concept of what it means, at an empirical level, to preserve an electronic record. In its efforts to figure out what must be done to preserve electronic records, the Preservation Task Force has recognized that, empirically, you cannot actually preserve an electronic record. You can only preserve the

---

Archival Perspective, 1997.

ability to reproduce the record. You cannot literally preserve an electronic record simply because it is impossible to store an electronic record as such. In the digital world, you can only store 'strings' or 'streams' of bits; that is, inscriptions on digital media that represent sequences of ones and zeros. A record, in contrast, represents, stands for, the acts and facts to which it attests. Between stored bit strings and electronic records there is therefore a very significant gap, a digital divide. In order to retrieve an electronic record, it is not enough to retrieve bit strings. In all cases, it is necessary to use hardware and apply software to translate the stored bits into the record. This situation is not unique to electronic records. A similar situation occurs in traditional media when records are encrypted for security. The encryption may contain the record, but unless we are able to decrypt it we do not have access to the record in any meaningful sense. Furthermore, if we do not apply the correct decryption algorithm, we may produce something that makes sense, be we cannot assert that it is the authentic record. The difference is that what is a special case in traditional media is the normal case with electronic records. The recognition that you can only preserve the ability to reproduce electronic records reinforces the viewpoint expressed in the *Guide to Managing Electronic Records from an Archival Perspective*. Not only is the purpose of preservation to enable continuing access to the records, but also the process of preservation includes retrieving and delivering them to users.

There are three general factors that compound the problem of reproducing authentic records from stored bits. The first is the ugly head of technology obsolescence. The second is that there has been very little proven success in overcoming obsolescence to deliver certifiably authentic electronic records across generations of technology. The third is that the production of electronic records is growing exponentially in volume and unpredictably in variety and complexity. To say the least, there is a great deal of uncertainty about the possibility of transmitting authentic electronic records over time.

There are many opportunities for errors in processing the stored bits to output a record, and in many cases there will remain, in spite of the best efforts possible, some measure of uncertainty as to whether the reproduced record is entirely the same as the original. A minimum condition for delivering authentic electronic records over time is to use the right hardware and the right software to reproduce the record, and to use them correctly. But even using the right hardware and software correctly does not necessarily eliminate all uncertainty. Consider a simple, narrative record, such as a one-page letter. Of course, we must preserve and deliver all of the content of the record intact; moreover, it is not unreasonable to expect that the content will be organized the way the creator produced it, and that the record will appear the same as it did to the creator. However, the appearance depends on many factors, such as the size and resolution of the output device used to display the record. In fact, the layout of the text can be changed significantly simply by changing the size of the window in

which it is displayed. Different software will produce different changes. For example, current generation software for *Netscape*, *WordPerfect*, and *MS Word* can all read and display files in HTML formats. If the file is opened in *WordPerfect*, and the size of the window reduced, the text will retain the same page, paragraph and line layouts it had in full screen, but to retain these attributes, *WordPerfect* decreases the size of the type. If the same file is opened with *Netscape*, reducing the window size will not affect page, or paragraph structure, or the size of type, but the flow of words across lines will change. *MS Word* preserves page, paragraph and line structure, and type size when window size is changed, but the result is that the user will only be able to see parts of lines. To read entire lines, the user will have to scroll back and forth in the window. Moreover, with many software products the appearance of a stored file changes considerably depending on whether the user chooses to look at it in print page format or in what is called “draft” or “normal" view. None of these examples involves any changes in the stored bits, neither those that solely represent the content of the record nor those that encode information about how that content is structured. The variability occurs only on the surface: at the interface between the computer and the human user.

Examples of the variability of the appearance of the same stored file used with a variety of different software applications, or even with the same application, could be extended much further. But the basic point is obvious: the appearance of an electronic record is not determined

solely by the bits that are stored. Do changes, such as different window size, different type size, etc., change the record itself? At an empirical level, clearly they do. But do such empirical changes have any significance in terms of the purpose the record serves, its ability to stand for the acts and facts it documents? If they do, then we would have to conclude that it is simply impossible to have an authentic electronic record because these problems of indeterminacy do not occur only in transmission of the record over time, or space, or technology. The appearance of the document can change on the same system practically instantaneously. At the other extreme, we could not categorically accept the conclusion that the authenticity of a record is independent of its appearance.

How then do we know - and on what basis can we certify - that when an electronic record is reproduced, perhaps years or even decades after it was created, it is authentic? One can describe three ways of determining the authenticity of a record. The first is from properties of the record itself or, in other words, on the face of the record. The second is from the archival context of the records; that is, from its provenance and the relationships it has with other records from the same source. The third method addresses the process of transmitting the record over time. The process model being developed by the Preservation Task Force accommodates all 3 methods of verifying the authenticity of records.

We can visualize the first two methods as putting a record into a time capsule. But how can you know that the record that comes out of the time capsule is the same as the one that went in? The first method establishes identity over time with respect to a defined set of properties of the record, such as the requirements for identity and integrity that were discussed in the papers of the Authenticity Task Force. The InterPARES Authenticity Task Force has identified a middle ground between these two extremes. It asserts that the authenticity of a record depends on its identity and its integrity. The identity constraint identifies specific attributes of all records, such as author, addressee, date, that must be known, controlled and invariant over time. The concept of integrity provides a tool that might be useful in resolving dilemmas such as those caused by the indeterminacy of the appearance of electronic records. It could be argued that the integrity of a textual record is protected when the appearance of the record is changed, if the changes impact the entire contents of the record in a uniform manner that does not alter the relationship among elements of the record. Obviously, the Preservation Task Force is looking forward to further refinements of these concepts from the Authenticity Task Force, but the preservation model being developed should be able to accommodate such refinements without basic changes.

The second method looks for evidence of authenticity in what Brent Lee calls *substantial similarity*. Rather than proving plagiary here, we want to determine how well any given record fits into the archival

fonds of the same creator. There are two facets to this, which we could call similarity and congruence. *A priori* we can assume that there will be similarities among records produced by the same creator, and the similarities will increase as we restrict our view to records produced in the same type of activity, and more so as we restrict the time frame and the documentary forms. Congruence refers to how a record fits into the sequence of records that is created in the same action or transaction. For example, an application for a research grant, the peer review of the application, and the grant award will have very different documentary forms, but they will share both content and the way the same elements of content are expressed in different records. Similarly, we would expect congruity between norms as to what data should be included in records and how that data should be expressed and the records that were created under those norms. We would expect congruity even in differences in the data: in other words, data related to the same subject or action should be expected to vary between prior and later records in accordance with the flow of the activity documented in the records. Both similarity and congruence provide circumstantial evidence of authenticity.

The third method of verifying authenticity is a logical extension of the archival concept of the unbroken chain of custody. We expand it to the idea of a chain of preservation. We can operationalize this concept by identifying all the links in the chain. Each link is an event or state in the lifecycle of the records that could affect the continued existence or authenticity of the record. Each is a potential point of

failure. If any link is broken, there are grounds for asserting, at the very least, that the authenticity of the record is questionable. And the burden of proof in such cases shifts to establishing that the problem associated with that link did not corrupt the record. In this perspective, the preservation process can be seen as a risk management method for avoiding, eliminating or counteracting known or knowable risks. This process must extend to every link. For this reason, the preservation model we are developing includes many activities that are not usually seen as part of preservation, from the accessioning of records into an archival system to the delivery of records to researchers. We are not modeling accessioning or reference and access functions. Rather, we are exploring where in those activities the continued existence or authenticity of the records is at risk, and we are designing a preservation process to mitigate those risks. To put it succinctly, in order to save the records we must be able to reproduce the authentic appearance, or sound, of the records themselves.

The Participation of Industry to the InterPARES Project*

Rich Lysakowski
Collaborative Electronic Notebook Systems Association, U.S.A.

Dr. Rich Lysakowski is Director of the Global Electronics Records Association (GERA) and Executive Director of the Collaborative Electronic Notebook Systems Association (CENSA), a market research and development association whose mission is to catalyze development of the markets for advanced automation products in all industries that are heavily dependent on research and development for success. CENSA has a strong focus on electronic records management and archiving systems.
Dr. Lysakowski also chairs the InterPARES Project's Global Industry Team, and is a member of the Preservation Task Force.

What I am going to discuss is the reason for the participation of industry to the InterPARES Project (and specifically to its Preservation Task Force), as well as how we see the role of the Preservation Task Force in the context of all the other outputs of the Project.

The current technology for high-value intellectual property records and regulatory compliance records is a high-speed, multi-media browsing and data management and record-keeping system. It is striking to observe the many parallels existing between the process of

* This paper has been transcribed from the audio recording of the

creating these systems and the creative process in the music field that Brent Lee explained earlier today. In the field of research and development, the overall shift to automation implies the utilization of literally hundreds of data sources in every imaginable format: in research it is not infrequent, for example, that new types of data are created in order to accommodate an altogether new research. We can have genetic sequences that are comparable to musical scores, or we can have actual audio and visual files that are captured in the laboratory. These are all included into today's automation systems -- the creative process of planning your experiments and executing them, collecting all the data, and putting everything in one place. Scientists are not interested in record-keeping, they just want to carry out their research and yet they're stuck, having to force their data into these systems, which are the only acceptable containers that our legal system understands and knows how to deal with. It is a serious problem.

CENSA is a global industry association. Our primary focus for the first three years has been on electronic records, their creation and their management. We used many results from the pre-project research conducted at UBC to feed into the requirements. We use the DOD 5015.2 as part of our requirements as well as FDA regulations and other regulations. To give you a sense of who is involved with CENSA and how it would contribute to InterPARES, I can tell you that CENSA represents some of the end users of the eventual project

---

Symposium's proceedings.

results -- 95% of our funds come from the end users because they are the ones who really want these problems solved. We also have supplier members, many of which will be familiar to you. One of our goals is to convince the suppliers to build the products that the end users specify. Some of the suppliers operate in the field of records management, some in the document management area, some deal with graphics, some others are laboratory vendors. They all have to fit together into an integrated system that produces documentary records.

You probably remember analogue data management and record-keeping systems: well, today we have things that are never analogue, because they are all *born* digital. Look at an electronic notebook: it is basically a document authoring tool that also performs data management. The data sources come from everywhere and go onto that page and what happens after that, becomes a record. Today we have these nice tools for assembling these compound documents but we still have to print them out to paper, sign the paper, witness the paper and put them into a traditional paper archive. The record becomes most important during patent infringements or interferences and also during year 18 out of a 20-year patent because that is the moment when usually a generic drug manufacturer will try to infringe on your patent. So you have to be able to retrieve your records 18 years after its creation.

So, we have all these great authoring tools that are coming on the market today, such e-books, the PDA's that some of you use, etc., but

they are all just data sources. They are not capable of keeping records, no matter what the vendors tell you. We are thus looking at this bifurcation on the back end, where whatever authoring tools you are using (the short-term re-use databases/data management systems) last anywhere from today to about five years, but when you go beyond that, things expire at different rates. All those complex compound documents expire at a different rate. Vendors go in and out of business or they drop a product line and you will not be able to access the data for reprocessing.

We also need a copy of the record for the legal system. There are static records, but we would also like to capture dynamic records, video, audio and the like. Still, we have this bifurcation to make things difficult, and we are relying on InterPARES to help us develop the long-term preservation and access.

Right now, we are under serious pressure from the industry to move fast. We have been pushing some of those urgencies into InterPARES to help it move faster too and deliver results that we can immediately put to use. Still, when we deal with electronic records, there are major risks, as you all know -- basically loss, loss, loss. We are afraid we are going to lose intellectual property and authenticity. IT museums are just not an option, but given that it would be impossible for all institutions to become IT museums, we need to solve the problem in a general way that is completely technology independent. In the United States, government regulations in this field include FDA, which was

the first agency, and every other agency including NARA, which followed shortly after. FDA came out with one of the first regulatory agency rules on electronic records and electronic signatures. We have GPEA, the Government Paperwork Elimination Act, which will force all agencies to go to fully electronic records by 2003. This just ripples down to industry. Agencies do not want to take paper records because they don't want to manage them. So what we need to do in InterPARES and everywhere else is to ensure that we have reduced the total risk for the electronic record to the smallest possible amount; we have to look at all the pieces of the system and at the people involved in programs and make sure that every one of them is reliable enough so that the sum total risk is about the same as we have with paper today. If we don't succeed, the legal system will not trust electronic records and will recommend keeping the papers -- something that we really cannot afford.. The pressures to get rid of paper are severe, as paper is too inefficient.

We realized early on that technology is not the answer, but a very small part of the answer. We came up with this concept of quality electronic records practices and programs and created the Global Electronics Records Association (GERA). This was initially a government-industry partnership but will eventually unfold into a full professional association. Initial focus is exactly on these quality practices; we have developed a model of what these practices entail. At the core is a reference model for electronic records programs and practices, but there are also other by-products such as the legal

acceptability guide for electronic records, which is actually the first one we are finishing. We came up with an unrefined, but fairly complete reference model, that we needed in order to produce the legal acceptability guide. This guide will be a public document available in June 2001[1]. Other documents are being developed over time. The reference model, in particular, needs the results of the InterPARES research in order to make it comprehensible.

We have identified the minimal required set of guidelines, reference models and standards in order to have a comprehensive program for electronic records. Among the elements that are needed we have a reference model that details what all the components are and a validation guide to make sure you have set up the system and program properly and that people are properly trained. Periodical auditing of the system will ensure that no hackers have gotten in and corrupted it, or that the trusted administrator is actually trustworthy. Then there are certification criteria and training materials. We are going to need some interim formats for electronic records collections. For the full reference model for the program, we are basically taking into account current laws and record-keeping practices. This drives policy, which in turn drives the mission and goals of the program and the record's organization, which again drives strategies and plans, roles and responsibilities. We need to have a reference model so that attorneys will be able to know whether they have all the necessary

[1] *The Legal Acceptability Guide for Electronic Records* was published by CENSA, Inc. in June 2001 (Editor's Note.)

and sufficient elements to be able to rely on those records in litigation.

From the point of view of the industry, the interest is to help us mature the product markets for electronic records quickly. We need what we call “whole products”, i.e. products that have all the essential features needed in a software and hardware system and that you can buy as a product with no need of purchasing along with it a huge systems integration engagement with a vendor. If this happens, what you end up with is a custom system, which you have to maintain yourself -- a very expensive proposition. We want therefore to mature these markets as quickly as possible, and to make sure that markets are healthy, i.e. that there are two or three strong competitors fighting for your business so that they keep each other honest and keep driving prices down.

We have been collaborating with the leading academic and national experts on archival science and records management, and have been part of InterPARES to ensure that the results can be applied in production systems within 2-5 years after the end of the project. Actually, we have already been applying the results. It is our goal to make sure that the commercial sector observes high standards of rigour. The IT industry itself has to be driven to embrace the concept of preservation, a concept that the industry has little awareness of. Today's model is rapid obsolescence, a model which of course doesn't address the needs of archivists. Among the other reasons behind our

participation in the Preservation Task Force is the fact that we have a large quantity and high levels of technology and engineering expertise, as well as vast experience with large-scale systems. We can help develop and validate rigorous models, process models, activity models, and the requirements to make sure they can be used throughout the public and private sector.

Talking about the private and public sector, it is obvious that ideally we would love to have a common model for government and industry: everyone would be using the same rulebook to operate their archives. It would also raise the level of professionalism throughout the world.

The results of the work of the Preservation Task Force will be reviewed by an extended industry team to ensure that all their needs are met. We will provide test beds for all the different results, the concepts, the models and the other tools. Generally speaking, the results that are coming from InterPARES are fairly high-level models that need to be driven down to an engineering level of detail that enables us to write codes from the models. We shall also help promote the results of InterPARES. The case studies, for example, are really important for us, as they validate all the models and the authenticity requirements. We have already conducted more than twenty case studies, although the models are not finished yet. In the industry's view, once the models are finished, we will need to conduct many empirical case studies to validate and refine them. We have to

make sure that all the Task Force results are integrated into a common set of models and that these models are completely technology independent.

Our expectations and contributions from the Preservation Task Force include a long list of adjectives. The models have to be general, detailed, predictive and extensible to different government and industry settings. They have to be empirically tested and validated. We shall have to "walk through" the models and use them to assess the quality of an organization's implementation of the models. Another reason why industry feels the InterPARES Project and its Preservation Task Force are important, is that better training is needed in industry. We hope that some of the talented people from the InterPARES Project will take jobs in industry because they have been intimately involved with the development. As you all know, it is not a matter of technology, but rather a paradigm shift from paper to electronic systems. We do not want to get rid of paper and do not feel comfortable with the idea. This is evidence of the fact that this is a cultural transformation, and a difficult one.

Using JARS to Preserve Electronic Records*

Bill Underwood
Georgia Tech Research Institute, U.S.A.

Bill Underwood is a Principal Research Scientist with the Georgia Tech Research Institute in Atlanta, Georgia. He is a member of the American InterPARES research team and a member of the Preservation Task Force.

I'm going to talk about a method of preserving electronic records that you see in JARS and I'll explain what JARS are. Computer scientists have developed communication protocols that are used for authenticating the sources of messages, verifying the integrity of messages and encrypting messages in order to maintain them secure. They analyse these protocols to determine whether they are correct and how they might fail, and in doing this they have developed so-called *logics of authentication* that are based on logics of belief, communication theory and set theory. Being equated with those, I saw the opportunity to combine some of the concepts, the axioms of those theories with the knowledge from the diplomatics and archival science, and then to apply them in an attempt to prove the correctness of some of our preservation methods.

---

* This paper has been transcribed from the audio recording of the Symposium's proceedings.

The InterPARES Authenticity Task Force has been formulating principles for determining the authenticity of electronic records. I have begun to formalize some of the authenticity requirements developed by the Task Force in the terms that are used in authenticating communications. One of the drafts of a definition that I have is: the digital record authentication is a type of identification whereby a party is identified as the author of the record, while the record's identity is provided by the name of creator, name of writer, name of addressee, chronological date, archival date, indication of matter/action and archival bond. In seeking to understand and reflect on this definition, the questions arise as to whether one is authenticating the chronological date, the name of addressee, etc., or whether those are not making up the identity of the record and one is authenticating the source. That would be more in line with the procedures used in communication theory and computer security.

From the InterPARES Draft Requirements for Authenticity we have this definition of data integrity that has been extended to digital record integrity: the property of a record whereby the content and form of the record have not been altered in an unauthorized manner since the time the record was created, transmitted or stored by an authorized source. One should note that I did not put the notion of integrity into the definition of authentication even though it is commonly associated with that. One of the reasons is that, with definitions such as these, it is possible to prove that if you were able to authenticate a message or record, the integrity of that message is

ensured. Now, that seems a little strange, but if you take into account the definition of authenticity and if indeed you do know the source of the record, and the time, you know the integrity. The reason for this is that if it had some other source, i.e. a forger, you would be able to discover it if you were sure of the record's authenticity. Still, you would not have the integrity of the record.

This is one of the technologies that I believe can be extended to support preserving the integrity of electronic records. It is not the only method, others are needed. But let's go back to JARS. The acronym JAR stands for *Java Archival Format*. It is a technology similar to *UNIX Tar*, and it combines many files into one. In addition to that, though, the technology was developed in such a way that the components of the file such as *Java* applets could be bundled into a single file with their data and quickly downloaded to a browser in a HTTP transaction. This format also allows you to just open one of the files and use it rather than having to open all of them. It also allows you to add individual files representing records to a JAR. JAR itself is an open industry standard.

One of the methods for preserving files representing records in a JAR consists in creating first a JAR file that contains the files of the collection and another file called a manifest file that contains the path and file names of the files. If a file representing a record is digitally signed, you will put that message digest for the file into the manifest file. Then, the file containing the public key certificate for the corresponding file is put into a signature file, so that we are

preserving the certificate as well. For those files that do not have a digital signature, a message digest is created to maintain their integrity and is put in the path file name also in the manifest file.

Now, associate what InterPARES has referred to as the identity of each file with its path file name in the manifest file. Then you create message digests for all of the message digests of the files, and also for the metadata representing the identity, that are stored with the message digests. This can either be in the signature file or the manifest file. Finally, sign the JAR using an archival private key and the message digest for the manifest file and insert the certificate for the archivists public key in the signature file.

Remember that the JAR itself is a singe file and if you were to view it in something like *WinZip*, you would see that there was a directory of meta-information that had a manifest file and in the same directory there was a signature directory that has certificates and other information. Here are the names of the files, but these would be the actual bit streams of the files.

I have extended the manifest file of the JAR to include, instead of key words and values, an XML style representation of the information from a record profile or information associated with the identity of the record. You will have a digest that has been computed for the message digests of all of the other files, and an entry for file where you can see the message digest for that particular file. One can represent the structure of a series and folders within that containing

file. Then, one can indicate, with attributes, the values representing the identity of the file as well as other information such as the format of the file that might be needed for viewing the file.

To verify the integrity of the preserved files, one extracts the files from the JAR, and to ensure the files from the JAR have not changed since the JAR was signed, the message digests of each of the records files in the JAR are recomputed and compared with the message digest in the manifest. Then the message digest in the manifest is recomputed and compared against the one that should have been stored in the signature file. What this does is ensure that no one has inserted a file or deleted one of the files. If so, the overall message digest will have changed. Furthermore, we have not changed any of the metadata. Then you would use the public key in the certificate in the signature file to verify that the digital signature applied to the manifest is that of a competent archival authority.

The general method that can be applied here to prove the correctness of the procedure is to express the assumptions of the communication protocols and the goal of the communication protocol in a logical language. Then, assertions are made in the logical language as to what is true after execution of each step. Then - and this is the big step - from the formalization of the concepts of diplomatics and archival science that are being used in these definitions, you should apply them, along with some axioms from set theory and axioms from a theory of belief, called "doxastic logic". Because what you're working with are presumptions of authenticity, you cannot know for an

absolute fact that it *is* authentic -- it is only from the evidence you have and from corroborating evidence, that you can reach your conclusions. Your assumptions amount to beliefs and even your conclusions amount to beliefs. Apply this theory to the assumptions and to the results of the steps to attempt to prove authenticity. I have attempted to prove that this method will indeed prove authenticity.

Here is a theorem that is proven (the proof is actually in the appendix to a paper that I presented at AIPA[1]) and that same paper has the necessary concepts and axioms that are necessary to prove the correctness. What is important are the qualifications that you put into the theorem. An important thing about this theorem is that it does not address the need for such things as refreshing media, conversions of records into new formats, or conversions of software to new hardware and software environments.

In proving theorems like this, you break the problem into sub-problems and prove what you can and then break the other methods into sub-problems and prove things there. Then you pull them all together to prove something like the correctness of an overall preservation method. The importance and significance of theorems such as this are that these methods can be used for active, semi-active

[1] AIPA is the acronym of *Autorita' per l'Informatica nella Pubblica Amministrazione* (Authority for Information Technology in the Public Administration), an Italian body set up in 1993 by a legislative decree with the specific goal to support and assist the different branches of the Public Administration in the gradual rationalization and simplification of the administrative activities (Editor's Note.)

and inactive records; while I only suspected that this was true, I found out that AIPA is actually implementing something similar for active records. That is, storing signatures and using an archival key with message digests for those files in order to verify the authenticity and integrity of records. Furthermore, the method can also be used during the transfers. So if you pull some records out of a records system that happen to be perhaps already in JAR or similar technology, and authenticate them and then repackage them in a JAR and transfer them, then the archivist after the transfer can open them up and verify that nothing has happened to them during the transfer to jeopardize their authenticity. Finally, when records are distributed to people who have requested them from an archive, you can verify the authenticity if you are able to verify the authenticity of the records that you have preserved, packaged up, signed and sent to a person. If you were able to attest to their authenticity, the individual on the other end can open them up and there can be procedures to verify the authenticity regardless of whether they have been unused for a month.

A General Narrative of the Preservation Task Force Models

Lisa Beitel, Deirdre Bryden
University of British Columbia, Canada

Lisa Beitel *is in her final term of the Master of Archival Studies programme at the School of Library, Archival and Information Studies of the University of British Columbia. Lisa completed her Bachelor of Arts degree in History at Simon Fraser University, Burnaby, B.C. She has been a research assistant for the InterPARES Project working on the Appraisal and Preservation Task Forces since September 2000. For the past three years, she has been working at the Vancouver Maritime Museum as Assistant Programme Coordinator.*

Deirdre Bryden *is in her final term of the Master of Archival Studies programme at the School of Library, Archival and Information Studies of the University of British Columbia. She has an Honours Bachelor of Arts in Archaeology and Classical Studies from Wilfrid Laurier University in Waterloo, Ontario. She has been a research assistant for the InterPARES Project working on the Preservation Task Force since September 2000.*

We are here to provide an overview of the models created by the Preservation Task Force. These models are based on the IDEF(0) methodology.

IDEF(0), otherwise known as Integration Definition Language, is a modeling technique for creating a structured graphical representation

of a system. It consists of a hierarchical series of diagrams, text, and glossary that are cross-referenced to each other.[1]

IDEF(0) is a simple and comprehensive modeling language that can be used to model a wide variety of activities. It is used by the Preservation Task Force because it provides a precise language that can be used to convey the intricacies of archival preservation.

IDEF(0) modeling demonstrates activities, which are represented by boxes. The objects and data relating to the activities are represented by arrows. When we look at the IDEF models created by the Preservation Task Force, we see a box that has four types of arrows connected to it. This box represents a function that is identified by a verb or phrase which describes the activity to be accomplished. Only one verb can be used to describe each activity.

The location of each arrow in the diagram has a particular meaning. Arrows going into the box from the top are called "control arrows". This class of arrows expresses the conditions required to produce the correct output.[2] In model A-0 we see *archival requirements; state of the art of information technology; and institutional requirements*. Entering the activity box from the left are "input arrows". This is the class of data or objects that are transformed by the activity into

[1] *Federal Information Processing Standard Publication 183: Integration Definition Language (IDEF0)*, (Gaithersburg: U.S.A National Institute of Standards and Technology, 1993), vii.
[2] *Standard Publication 183*, 4.

outputs.[3] For example, the inputs for this model are: *information about electronic records selected for preservation; transfer of electronic records selected for preservation; and request for records and/or information about records.* Exiting the activity box from the right, are the "output arrows". This represents the data or objects produced by the activity.[4] In this model, the outputs are: *requested records or record components; certificate of authenticity; information about preserved records; and information about preservation.* The "mechanism arrows" enter the activity from the bottom of the box. Mechanisms are the means used to perform an activity.[5] For this model, the mechanisms are: *digital preservation technology, facilities, and persons responsible for preservation.*

This model, A-0, is the top-level diagram which provides the most general description of the preservation process. This diagram, known as a "parent model", is followed by a series of child diagrams which provide a hierarchical decomposition of the specific activities involved.

The first level of the child diagrams is A0 "Preserve Electronic Records." It is decomposed into 3 primary activities; "Manage the Preservation Function", "Bring in Electronic Records" and "Maintain Electronic Records."

---

[3] *Standard Publication 183, 5.*
[4] *Standard Publication 183, 6.*
[5] *Standard Publication 183, 5.*

Let's begin at A1. This activity "Manage the Preservation Function" is where the decision is made about the methods and strategies to be used to preserve electronic records. There are three controls on "Manage the Preservation Function". The *archival requirements* control is based on archival science and diplomatics, along with the standards and best practices current in the archival community. Another control is the *state of the art of information technology,* which is related to technology's ability to satisfy archival preservation requirements. It includes not only the products available in the market place, but also the knowledge and possibilities available from computer science and engineering and related disciplines. The final control is *institutional requirement*s. This control involves the requirements imposed on the institution. It is assumed that the institution will do its own analysis and establish its own institutional requirements such as legal, societal and cultural limitations. Institutional requirements can be both internal and external.

There is only one mechanism affecting A1, which is *persons responsible for preservation.* This is self-explanatory, for it refers to the persons authorized to carry out the preservation function.

One of the inputs to A1, *information about electronic records selected for preservation*, is a product of the appraisal process in that it identifies and characterizes records which are to be preserved. This information is needed in order to plan for preserving these records.

Another input into A1 is *information about transferred electronic records*. This input encompasses archival and technical information about the transferred records, such as provenance and the format of bit files when they are actually received in the archives. It is used to compare with the information from appraisal and to adjust plans, if needed.

The final input into A1 is *management information about preservation*. This information is an outcome of the activities of A2 and A3. When these activities take place, they produce specific information about the maintenance and reproduction of the record. As an input, this information enables management to update their preservation strategies and to redefine objectives and targets.

One of the outputs of A1 is *information about preservation* which includes reports that account for the preservation of the records. This information serves as an audit report, on the assumption that the person responsible for carrying out preservation is accountable to a higher authority.

Another output of A1 is *preservation strategies*. These are methods used for maintaining components of electronic records and their related information, both over time and in authentic original order, along with objectives and targets that management establishes for applying these methods.

The final output for A1 is *transfer control information*. This is a very significant output because it employs the specifications and criteria necessary to ensure that the records transferred are actually those selected in appraisal. For example, it details when the records selected for preservation should be transferred to the archival system, or who has the authority to transfer the records.

The next activity, A2 " Bring in Electronic Records", is when the records selected for preservation are brought into the custody of the preserver and are formally accessioned for long-term preservation.

Two of the controls, *preservation strategies and transfer control information*, are direct outputs from A1. The information sets the specific conditions that are required to ensure the proper accessioning of records. The preservation strategies determine if any special processing, such as migration of bit file formats, is needed to preserve the records.

Another control, *accessioning policy*, sets institutional policies and individual responsibility over the records transferred for preservation.

A2 has the same mechanism as A1, which is *persons responsible for preservation* (the persons authorized to carry out the preservation function.) It also uses information technology to process the records, but that is not shown in the diagram because, in the parent diagram

you saw, it is indicated that technology is a mechanism in all child processes.

There is only one input for A2, which is *transfer of electronic records selected for preservation*. This transfer involves both the selected records themselves and the information about them.

There are two outputs, *electronic records accession, and management information about preservation*. The *electronic records accession* is those records successfully accepted for preservation, along with relevant information needed to preserve and reproduce the records.

As previously discussed, the *management information about preservation* returns as a feedback into the A1 activity.

The next activity is A3, "Maintain Electronic Records". This activity manages information about records, maintains the ability to retrieve and reproduce records, and actually reproduces records, or outputs their digital components when needed. Like A2, *preservation strategies* serve as a control on this activity, and *persons responsible for preservation* act as the mechanism.

One of the inputs, *information required to certify authenticity* carries information about the integrity, adequacy, and correct execution of the preservation function.

As an output of A2, *electronic records accession* becomes another input into the A3 activity. As previously stated, the records are those successfully accepted for preservation, along with the relevant information needed to preserve and reproduce the records.

The final input into A3 is *request for records and/or information about records*. This is a request to generate preserved records. In some instances, there may be a request only for the information about records. For example, if that information is not available through archival descriptions or finding aids, but only from the audit trail.

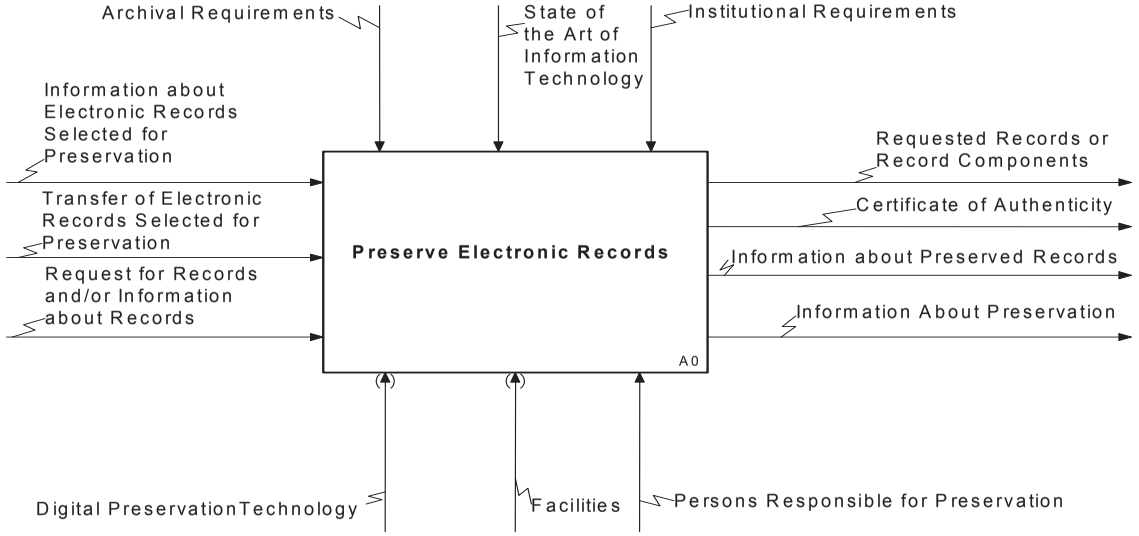
An output from A3 is *information about preserved records*. This is the specific information asked for in the request for records and/or information about the records.

Another output is *requested records or record components*, which is comprised of the reproduced records or record components and the information about how to reproduce them in another system.

The final output is the *certificate of authenticity.* This output is the attestation by the person responsible for preservation that one or more of the requested records are authentic.

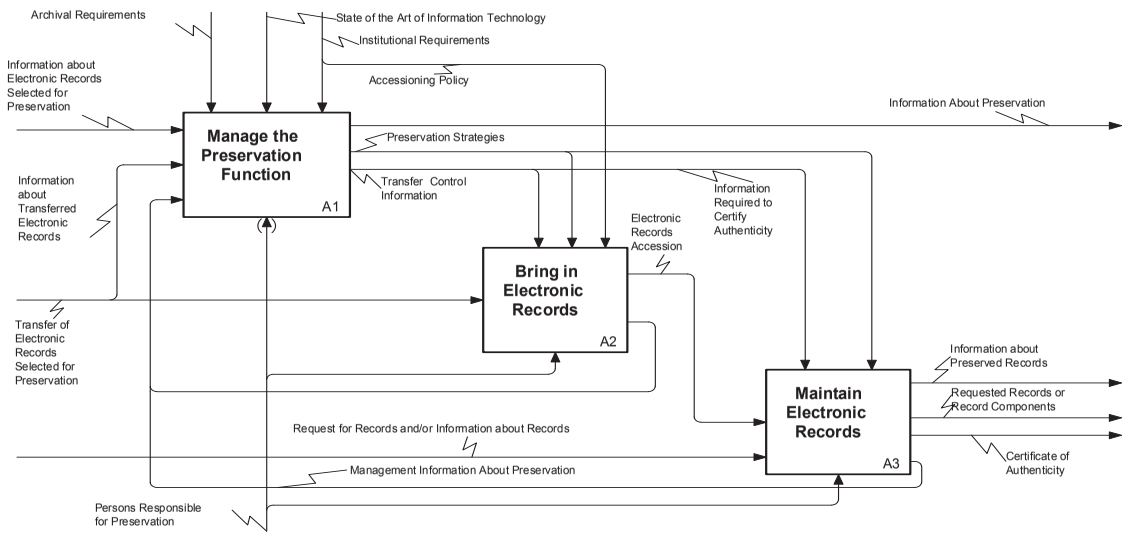
In conclusion, the whole purpose of the preservation modeling exercise is to determine how to maintain the ability to reproduce authentic electronic records.

USED AT:	AUTHOR: Preservation Task Force	DATE: 2/16/2000	■ WORKING	READER DATE	CONTEXT:
Vancouver Workshop #7	PROJECT: InterPARES Project	REV: 2/16/2001	DRAFT		TOP
	NOTES: 1 2 3 4 5 6 7 8 9 10		RECOMMENDED		
			PUBLICATION		



NODE:	TITLE:	NUMBER:
A-0	Preserve Electronic Records	4.0

USED AT:	AUTHOR: Preservation Task Force	DATE: 2/17/2000	■	WORKING	READER DATE	CONTEXT:
Vancouver Workshop #7	PROJECT: InterPARES Project	REV: 2/15/2001		DRAFT		[redacted]
				RECOMMENDED		
	NOTES: 1 2 3 4 5 6 7 8 9 10			PUBLICATION		A-0



NODE:	TITLE:	NUMBER:
A0	Preserve Electronic Records	v 4.0

An Overview of the InterPARES Glossary

Philip B. Eppard
State University of New York at Albany, United States

Philip B. Eppard is Dean of the School of Information Science and Policy at the State University of New York at Albany, where he is also an associate professor teaching classes in archives, preservation management, and rare books. He has been editor of the American Archivist *since 1996. He is director of the American InterPARES research team, a member of the Appraisal Task Force and a member of the Glossary Committee.*
Before joining the faculty at Albany in 1988, Dr. Eppard was the archivist for the Archdiocese of Boston. He has written and edited articles and books in the areas of American literary history, bibliography, documentary editing, and archives. He has a doctorate in American civilization from Brown University.

One of the great challenges of a multi-national, multi-disciplinary project such as InterPARES is ensuring that the terms and phrases used in the research are used consistently within the project and that their meanings are communicated effectively to the world outside the project. The InterPARES glossary is being created in order to achieve these goals. Its officially stated purpose is "to facilitate the communication of ideas and research findings by clearly defining key terminology that is used within the InterPARES Project." In one sense, the glossary could be regarded as a byproduct of the main thrust of the research of the project into the theory and methods underlying the permanent preservation of authentic electronic records. The definitions that make up the glossary are, after all, drawn out of

the work of the task forces and teams, created, as it were, in the normal course of the business of the research. The reality, however, is that the glossary will likely be one of the most important products of the research.

The language of electronic records is a problem that inhibits both progress in improving our abilities to manage and preserve electronic records and progress in electronic records research. The heart of the language problem lies in the variations in meaning between the language of records creators, information technology personnel, and archivists and records managers. These differences have been made readily manifest to InterPARES researchers carrying out case studies of electronic records in their interviews with the different people involved in creating electronic records. Records creators have been quick to ask researchers some very fundamental questions about definitions, the most basic being: "What do you mean by a record?" Database administrators, for example, will talk about records in ways quite different from archivists. Furthermore, many of the terms used in InterPARES have their roots in diplomatics, a discipline with which most archivists and records managers, in North America at least, are unfamiliar. The meaning and usage of these terms in diplomatics are even more alien to records creators and information technology professionals. Such problems of variation in meaning and usage are only compounded when dealing with different countries that have their own varied usages, which can sometimes differ even from region to region within a particular country.

The first public version of the InterPARES Glossary was released on the InterPARES web site in fall 2000. This edition includes eighty-two terms, many of them drawn from the work of the Authenticity Task Force. These are primarily terms used in the Template for Analysis, and so they include a fair number of terms from diplomatics. The glossary is described as "an alphabetical list of terms along with their definitions, scope notes, and equivalence relationships." It is important to note that it is the *InterPARES* glossary and that its goal is not to be a comprehensive glossary covering all archival terminology, not even all electronic records terminology. Yet given the scope of the InterPARES project, the scope of the glossary will certainly be wide, and we anticipate that the glossary will have valuable applications beyond the task of facilitating an understanding of the work of the project.

The responsibility for overseeing the creation of the InterPARES Glossary has been vested in the Glossary Committee. While the principle work of InterPARES is organized around the four task forces of authenticity, appraisal, preservation, and strategies, the Glossary Committee of necessity crosses the boundaries of the four task forces. It is composed of one representative from each of the task forces. Ken Hannigan of the National Archives of Ireland is the chair of the committee, and Torbjörn Hörnfeldt from the National Archives of Sweden serves as vice-chair.

The project has established a clear procedure for development of the glossary. Terms used in any of several places in the InterPARES research may be nominated for inclusion in the glossary. That means terms used in meetings, workshops, forum discussions, models and other products of the research, as well as supporting literature used in the research may become entries in the glossary. Nominations must meet at least one of seven stated criteria for inclusion in the glossary. Some examples of the criteria are:

- The term represents an entity or concept that is key to understanding the research questions or findings of an InterPARES research domain.
- The term has conflicting definitions and or/meanings between the disciplines or sector.
- The term has multiple meanings.
- The term is an obscure word or phrase that the majority of InterPARES researchers are not expected to be familiar with.

Most glossary terms will be proposed by one of the Task Forces, and nominations will flow from the Task Force to the Glossary Committee via the Task Force's representative on the Committee, but they can also be nominated by members of the international research team or by members of the Glossary Committee itself.

The approval of glossary terms is a multi-staged process. First, nominated terms are presented to the Glossary Committee, and the

committee either accepts or rejects the term for inclusion in the glossary. In order to propose a term, an InterPARES researcher completes a new term proposal form, which includes the name of the term, a proposed definition, and source for the definition, the context in which the term is used in the project, and how the term complies with stated nomination criteria.

The Glossary Committee reviews new term proposals using a "quick check" voting method under which only dissenting votes are cast. New terms are posted to the Glossary Committee Forum on the InterPARES web site, and members of the committee have ten days in which to register a dissenting vote. If no dissenting votes are cast, the term is accepted and then goes to the graduate research assistants at the University of British Columbia for further research. Terms for which a dissenting vote has been cast will be discussed at the next meeting of the Glossary Committee for resolution.

The research assistants' review of the nominated terms will include compiling other definitions and usages noted in the literature. It will also include a check of instances where there may be potential conflict in usage between terms nominated by different task forces. Lists of nominated terms with all the appropriate documentation will then be presented to the Glossary Committee on the last day of each month. Once again the "quick check" voting process will be used. Approved terms will enter the full glossary immediately, while those

about which there is dissent will be discussed at the next meeting of the committee.

The InterPARES glossary is an ever growing and evolving entity. New terms can always be added, of course, but existing terms and definitions may also need to be revised during the course of the project. Therefore there is also a formal process for submission and approval of revisions to glossary terms. Revision requests will be given to the appropriate Glossary Committee member, who will determine the validity of the request and post it to the Glossary Committee Forum. Research assistants will then research what the impact of the revision might be on other terms in the glossary. Once again, the "quick check" voting procedure will be used to approve revisions to glossary terms.

The current InterPARES Glossary will grow significantly over the next several months as the various task forces move toward the conclusion of their work. It is worth noting also that while the public glossary will simply present the alphabetical list of terms and definitions, the glossary system which is managing the process, will include a rich set of data on term histories, usage notes, and research on alternative definitions. Finally, the glossary will be translated into French, as is appropriate for the product of a Canadian research project. Translation of the glossary into other languages in the future would also be in the spirit of the InterPARES Project and help

facilitate the goal of enhancing international understanding of its research in particular and the world of electronic records in general.

The InterPARES Glossary System*

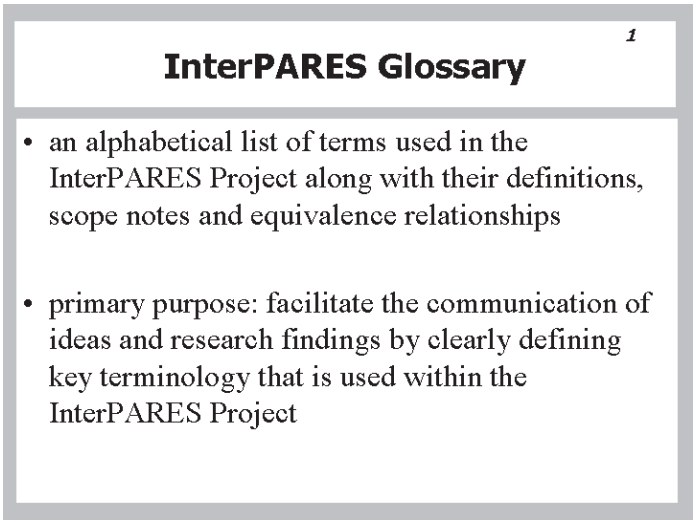
Peter Van Garderen
University of British Columbia, Canada

Peter Van Garderen served as Project and Technical Coordinator for the initial two years of the InterPARES Project (1998-2000). He is an adjunct professor at the School of Library, Archival and Information Science at the University of British Columbia. Peter has also been a software product manager and is currently employed as System Architect Consultant for his company, Artefactual Systems, Inc.

The InterPARES glossary system is essentially a by-product, not a direct product of the research itself. Philip Eppard gave you a good overview as to what the InterPARES glossary is and its purposes. I am sure you can see that there is a great deal of information management involved with putting the glossary together, but there are also very many credibility issues as far as how the terms get nominated and accepted into the glossary, considering that we are working with such a large, multi-disciplinary group of people from different theoretical and national backgrounds. We want this to be a very open process, especially considering that it is our common hope that the legacy of the InterPARES project will be a long one, and that there will be a lot of evidence and rationale for how we came up with certain terminology and how we feel it fits into the rest of the

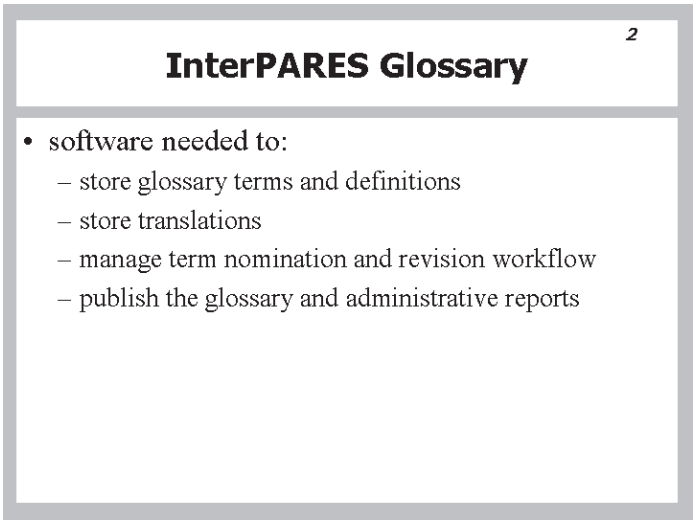
* This paper has been transcribed from the audio recording of the

terminology that deals with the issues we deal with.

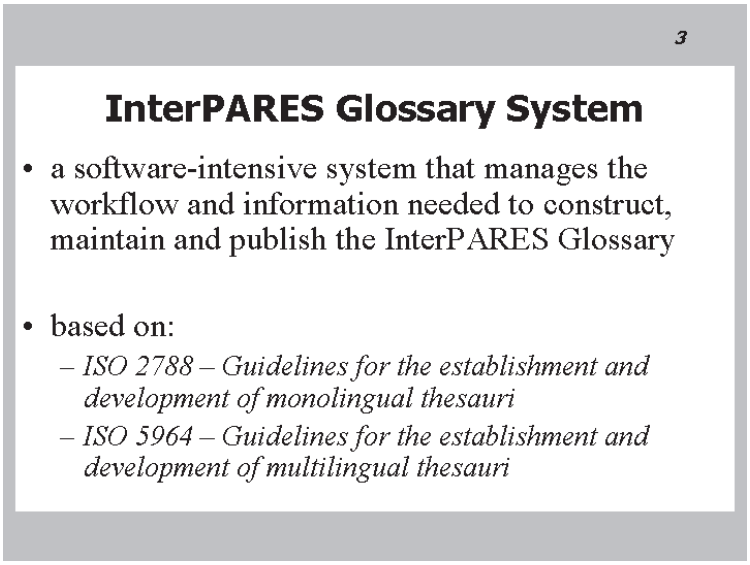


When we decided to put together a glossary, it became fairly obvious that the best way to approach the challenge was with some specific software. I did a market analysis of the existing controlled vocabulary software and found that the majority of them could not deal with any of our major top requirements that you can see in slide number 2.

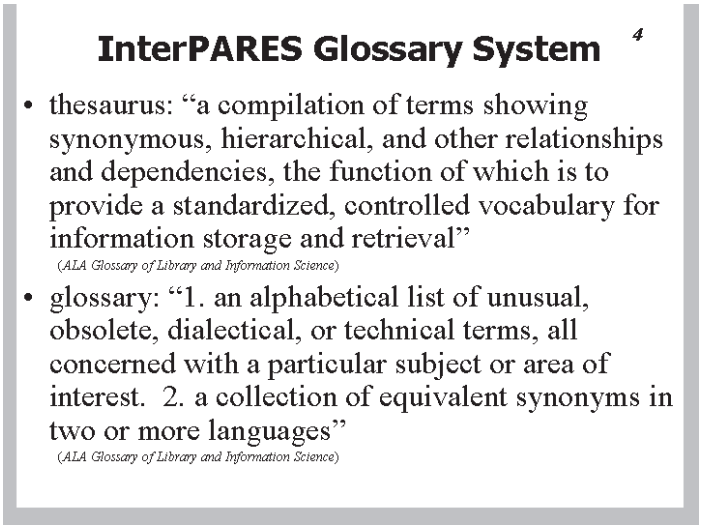
Symposium's proceedings.



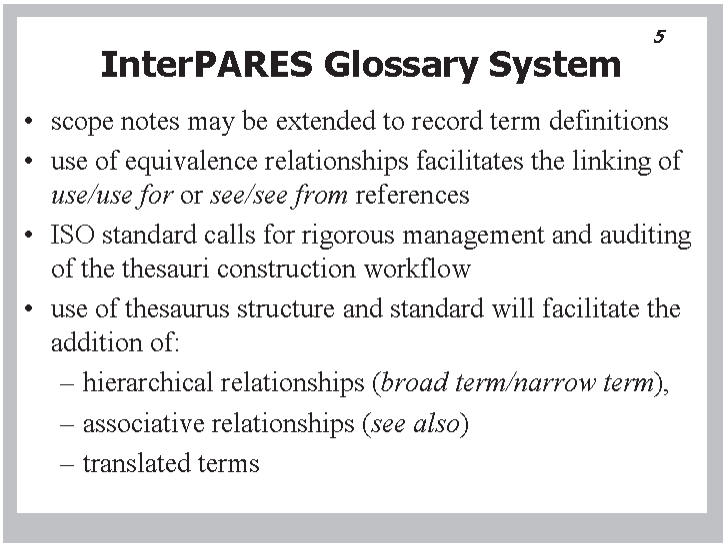
In particular, there was a problem with handling the fairly detailed workflow that we were anticipating as well as the multi-lingual aspects of the system. We decided to embark on our own development project to put together an InterPARES glossary system -- in this case, a software intensive system, implying that software is only part of the system as obviously the workflow and the people in the system are important parts as well.



The applicable standard upon which to base the glossary turned out to be the ISO standards for thesauri construction. But why a thesaurus? Well because the main function of a thesaurus is to assist indexing and retrieval. A glossary, as we heard earlier, is an alphabetical list of terms and their definitions and it helps to manage the equivalent relationship of synonyms amongst terms. However, there is no standard for glossary construction. There are some American standards, but the ISO thesaurus, being an international standard, offered better adaptability to our purpose -- a part, as it is, of an international project.



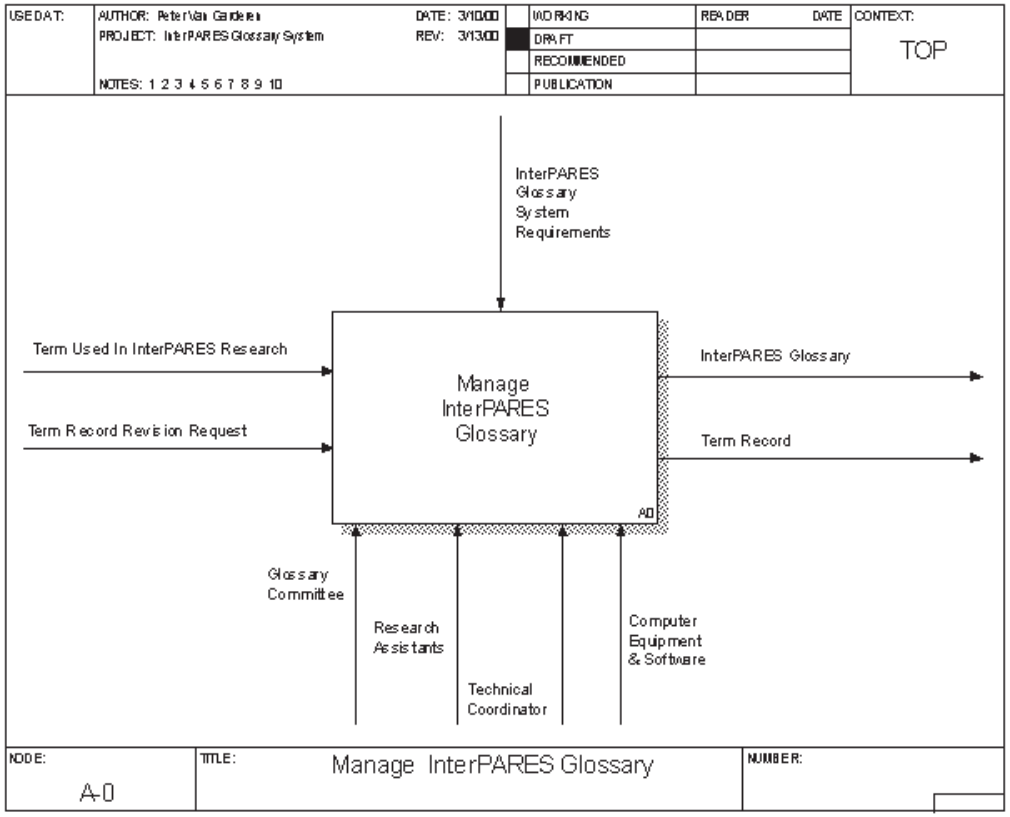
The actual structure of the thesauri, as well as the underlying glossary, helped us first of all to manage our equivalence relationships. The standard allows for expansion of scope notes to include term record definitions. The ISO standard also calls for and supports the rigorous management and auditing of the thesauri construction workflow. We are anticipating in the future that the glossary will have to handle hierarchical relationships, especially as we get to lower levels of detail, associative relationships, and of course translated terms, all of which the ISO standard help us incorporate.



The idea was to put together a sound design at the bottom as the glossary expanded and the needs of the glossary expanded, so that the design and the system could expand as well. In doing this, I have found that the ISO thesauri standard is in fact a good standard exactly for doing that. It is able to output a number of different types of controlled vocabularies, including glossaries, classification systems and name authority files.

Based on the directions of the Glossary Committee and the ISO standard, I embarked on IDEF(0). I did a functional decomposition of the glossary system process, which gets into much more detail than you can see in slide number 6. Then the functional decomposition was broken down into detailed workflow procedures, which became then the basis for the actual written procedures that the glossary committee follows now. At the end, based on the requirements

documentation and the system design, an actual software database and user interface software was developed at UBC and the existing components of the InterPARES Glossary system came together.


USED AT:
AUTHOR: Peter Van Garderen
DATE: 3/10/00
PROJECT: InterPARES Glossary System
REV: 3/13/00
NOTES: 1 2 3 4 5 6 7 8 9 10
WORKING
DRAFT
RECOMMENDED
PUBLICATION
READER
DATE
CONTEXT:
TOP
InterPARES Glossary System Requirements
Term Used In InterPARES Research
Term Record Revision Request
Manage InterPARES Glossary
A0
InterPARES Glossary
Term Record
Glossary Committee
Research Assistants
Technical Coordinator
Computer Equipment & Software
NODE:
A-0
TITLE:
Manage InterPARES Glossary
NUMBER:

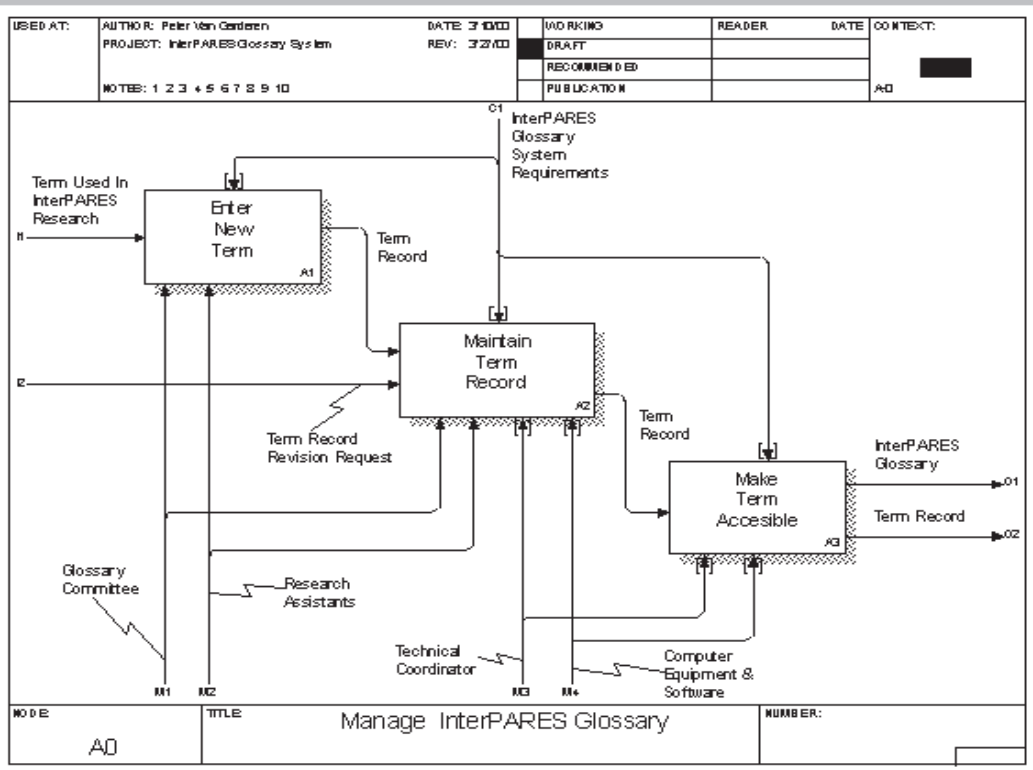

USED AT:
AUTHOR: Peter Van Garderen
DATE: 3/10/00
PROJECT: InterPARES Glossary System
REV: 3/22/00
NOTES: 1 2 3 4 5 6 7 8 9 10
WORKING
DRAFT
RECOMMENDED
PUBLICATION
READER
DATE
CONTEXT:
A-0
C1
InterPARES Glossary System Requirements
Term Used In InterPARES Research
I1
Enter New Term
A1
Term Record
Maintain Term Record
A2
I2
Term Record Revision Request
Term Record
Make Term Accesible
A3
InterPARES Glossary
O1
Term Record
O2
Glossary Committee
Research Assistants
Technical Coordinator
Computer Equipment & Software
M1
M2
M3
M4
NODE:
A0
TITLE:
Manage InterPARES Glossary
NUMBER:

Diagram: A11(a)- Nominate Term
Model: InterPARES Glossary Workflow

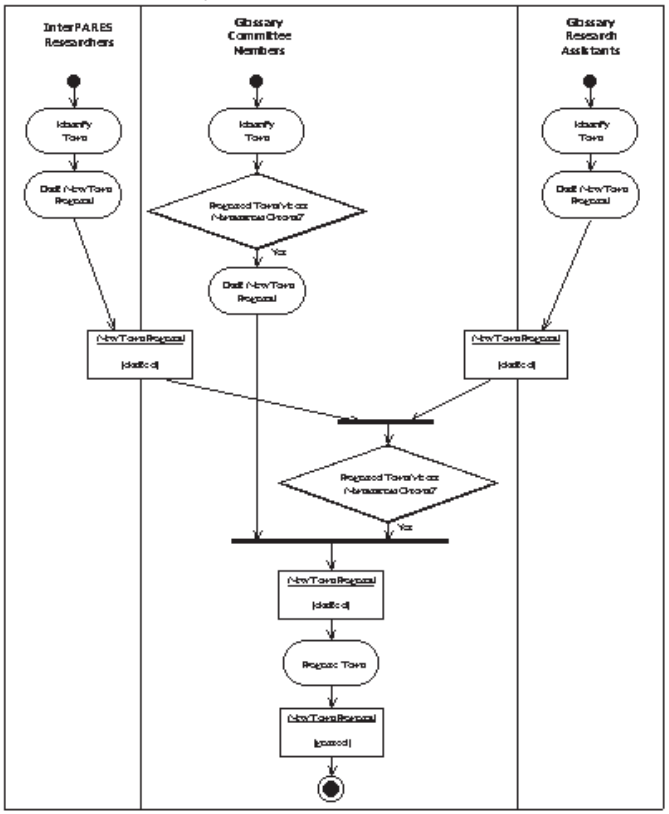
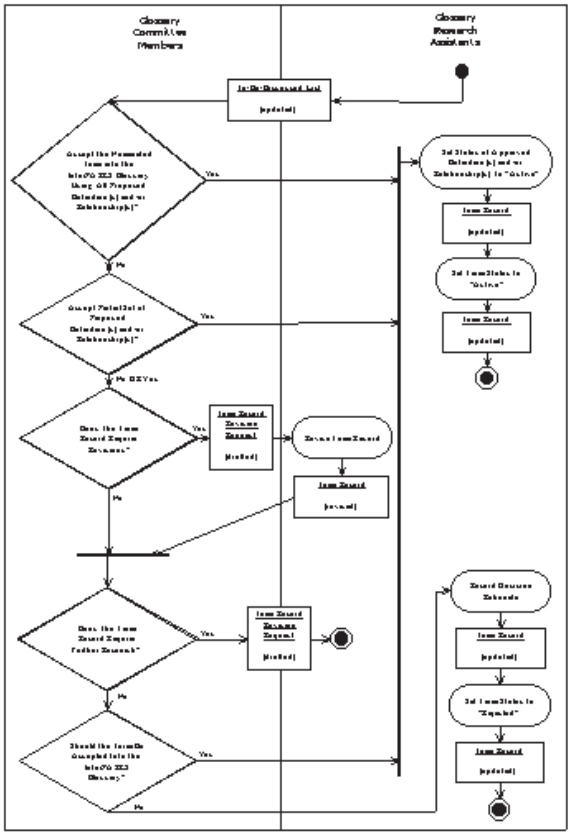
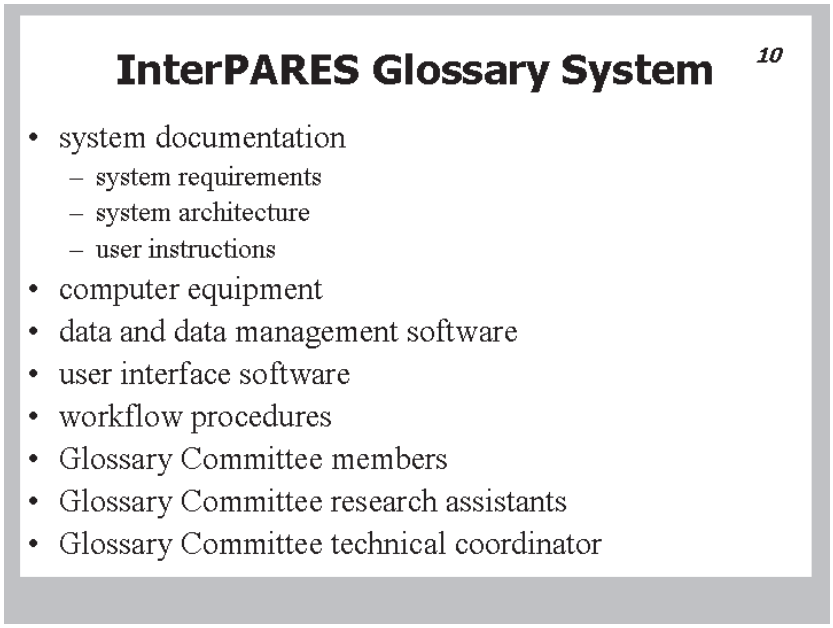


Diagram : A13(b) - Determine Term Status
Model: InterPARES Glossary Workflow



I would like to stress once again that what we have developed is not a piece of software. The software plays an important role, the software stores the data and helps the research assistants manage the workflow procedures, but obviously the system also includes documentation and the actual people who carry out the different procedures.



The main entities within the system are term records which store all the information for the individual terms, their definitions, their equivalent relationships, the place within the InterPARES documentation where the term was used, links to citations for alternative definitions, and the ability to mark active, inactive or nominated definitions -- in other words, all the contextual

information that is needed to manage the individual terms and provide a history of their use as well as their relationship to other terms and term definitions. The new term proposal is another entity that is used to actually propose another new term by glossary committee members and a term record revision request is the entity that is used to make request to changes to the term or its term definition. The system allows us to output a number of reports to support that workflow, including the very technically named "to be discussed list".

The user interface system and the database is a simple Access database at this point in time. The Term Records Forms and their tabs allows us to capture definitions and record more detailed information about the source citation of the individual term definitions, as well as manage the term relationships and the actual term administration. At this point there is no audit trail yet of the different revision requests that might have happened to this term, although these would be indicated. The plans are to update this prototype and to make the user interfaces web accessible, so we can better integrate the workflow with the database management system.

The Development of Style Guidelines for the InterPARES Project Glossary

Robert Edwards
University of British Columbia, Canada

Robert Edwards has degrees in law and history from the University of Alberta. Before entering the Masters of Archival Studies program at the University of British Columbia in 1999, he worked as a free lance writer and editor. He taught English as a second language and briefly, some years ago, he practiced labour law. As a research assistant for the InterPARES Project, he has drafted style guidelines for glossary terms and definitions.

One of my responsibilities as a research assistant with the InterPARES Project has been to produce a set of style guidelines for use in the construction of the glossary.

The intention of the InterPARES Glossary has been to construct, maintain, and make accessible "a multi-lingual controlled vocabulary consisting of terms used in the InterPARES Project"; we want to provide definitions for terms used in the Project so as to promote efficient communication of concepts, both by the participant researchers as the Project develops, and by other parties who would like to make use of its data and findings. Thus, dissemination of the information that is produced is an important consideration.

By referring to it as a controlled vocabulary we have meant that we wish, first and foremost, to produce a set of terms and definitions, drawn from the several different domains of research engaged in by the project, that are as much as possible conceptually consistent.

The achievement of this aim will be facilitated by applying a degree of standardization of literary style used for drafting definitions. To this end, we have surveyed many existing glossaries, in various fields and disciplines, and based on that research have tried to lay down, if you will, our own house style.

This has included such elements as:

spelling (we have decided to use the Oxford English Dictionary, seeing it, as we do, as possessing a certain authority as an international standard);

the use of definite and indefinite articles;

the use of complete sentences or sentence fragments;

punctuation and capitalization.

Since information retrieval will be an important part of dissemination of InterPARES research - and to this end, the construction of a thesaurus will be an important contribution -, we have decided that we would attempt to :

a) track the relationships between terms in the glossary - equivalent terms, terms related hierarchically, and, to some extent, associated terms or concepts; and

b) apply a standard to the form of terms that appear in the glossary, according to the International Standardization Organization's guidelines for thesaurus construction.

At present, though, we are not constructing a thesaurus, which is a different thing than a glossary -- its purpose is to promote information retrieval. A thesaurus will eventually be constructed, and by performing our research with the needs of a thesaurus in mind, we will, through the glossary system, compile much data that will be a valuable resource for that task when it is eventually undertaken.

Because, first of all, as I have said, a glossary and a thesaurus are different and serve different functions, we have had some difficulty in applying the term form guidelines. The relationship between the form of term used - plural vs. singular, grammatical form of compound terms, for example - has been somewhat uneasy.

We have decided that the term form appearing in the glossary will largely be based on usage by the Task Force that submits the term and definition. We are acknowledging that glossaries do not appear to follow the thesaurus guidelines for term form. So, we can see this one of two ways: either we are interpreting the guidelines so as to allow us more latitude; or, for the front end of the glossary, if you will, the guidelines need not be followed.

However, insofar as we are tracking in our research the data available on usage by the various professions and disciplines involved or

interested in the project, we are creating a resource that will be used at the thesaurus construction stage.

_____

The InterPARES Project Strategies Task Force: Preliminary Report

Sharon E. Farb
University of California, Los Angeles, U.S.A.

Sharon E. Farb coordinates the acquisitions of digital products for the UCLA library and advises the library on legal and policy issues including intellectual property. Sharon is a member of the American Team and a member of the Strategies Task Force.

Introduction

The International Research on Permanent Authentic Records in Electronic Systems (InterPARES) project consists of four primary research domains. A Task Force of international experts in the field represents each research domain. In October 2000, Luciana Duranti, Project Director, convened the Strategies Task Force. The work of the Strategies Task Force builds on the work of the three other task forces. The Authenticity Task Force (domain one) studies the requirements for preserving authentic electronic records over time. The Appraisal Task Force (domain two) determines the criteria and methods for selection of authentic electronic records over time. The Preservation Task Force (domain three) develops methods, procedures and standards for the preservation of authentic electronic records over time. Building on the work of the Authenticity, Appraisal and Preservation Task Forces, the Strategies Task Force

(domain four) formulates guidelines to develop international, national and organizational policies, methodologies, and theoretical frameworks related to the authenticity and preservation of electronic records over time. According to Lewis Branscomb and Brian Kahin, prominent policy researchers, development of a theoretical framework and standards provide for more coherent policy development, contribute to the development of models of best practices and establish a process for addressing interoperability, harmonization and improvement of cost efficiencies.[1] Branscomb argues that, in the absence of a coherent intellectual framework, conflicts arise because traditionally distinct domains and multiple approaches result in inconsistency, lack of interoperability, permanent loss of the cultural, historical and financial record and the inability to implement long-term strategies and goals.[2] The goal of the Strategies Task Force to develop a theoretical framework that can be used to guide the development of policies, methods and standard for selecting and preserving authentic records over time is, therefore, a critical component of the InterPARES Project.

This paper will first present an overview of the InterPARES Project highlighting the work of the Task Forces. Secondly, it will describe the research plan and objectives of the Strategies Task Force

[1] Branscomb, Lewis M., and Kahin, Brian. "Standards Processes and Objectives for the National Information Infrastructure" In *Standards Policy for Information Infrastructure* edited by Brian Kahin and Janet Abbate. (Cambridge, Massachusetts: MIT Press, 1995).
[2] See footnote 1 above.

including key research questions and deliverables. Thirdly, it outlines the Task Force's methodology, timeline and action plan. The final section highlights future research.

Project Overview

Luciana Duranti, Professor in the Master of Archival Studies Program at the School of Library, Archival and Information Studies of the University of British Columbia, founded and directs the InterPARES research Project. The focus of the InterPARES research, and of today's symposium, is on the development of strategies, methods and theoretical frameworks to guarantee the long-term preservation as well as the authenticity of electronic records. As noted by Luciana Duranti, the last decade has generated more recorded information, much of it electronically, than any previous decade. The majority of this information, however, is less accessible today than ever before and even less likely to be accessible to future generations.[3]

There are several concurrent obstacles to developing long-term preservation of authentic electronic records that InterPARES researchers are actively addressing. The first obstacle is the exponential growth of information technologies used to generate, manage, store and provide access to digital records. For example, in one large United States federal department such as Health and Human

[3] Luciana Duranti, InterPARES Project Director, Presentation to International Symposium entitled *How Do You Know it's the Real Thing?: Authentic Records in the Electronic Age* (Istituto Italiano di Cultura, Vancouver and the InterPARES Project), University of British Columbia,

Services "an estimated one million email messages are now exchanged each day."[4] Second, information technologies are often proprietary and quickly superceded resulting in incompatibility, lack of interoperability and technological obsolescence. Third, evolving digital multimedia technologies as well as numerous hybrid systems involving digital and non-digital records require new approaches and solutions to insure the authenticity and preservation of digital information objects in context over time.

The Strategies Task Force Research Plan

The primary objective of the Strategies Task Force is to develop an intellectual framework for the translation of research results into international standards and national and organizational policies and strategies. The Strategies Task Force identified three key research questions to guide the research of the Task Force. First, what policies, strategies and standards will protect the authenticity and preservation of electronic records over time? Second, what are the principles that can guide the development of international strategies, standards and policies for the long-term preservation of electronic records? Third, what are the criteria for developing national and organizational

---

Vancouver. February 19, 2000.

[4] Jason Baron, Esq., Presentation entitled "E-Mail Litigation Wars: the U.S. National Archivist Strike Back*"* given at the International Symposium *How Do You Know it's the Real Thing?: Authentic Records in the Electronic Age* (Istituto Italiano di Cultura, Vancouver and the InterPARES Project) University of British Columbia, Vancouver, February 19, 2000.

policies, standards and strategies for the authenticity and long-term preservation of electronic records?

Collaboration is the Key

Recognizing the importance of collaboration across multiple research domains and disciplines, the Strategies Task Force comprises of the Project Director, the Chairs of Authenticity, Appraisal and Preservation Task Forces as well as two experts in legal and information policy issues. The composition of the Task Force reflects the dual importance of broad-based collaboration across disciplines and domains as well as the corollary of synthesis and integration of research into practice.

The principal deliverable of the Strategies Task Force is the development of an intellectual framework that can be used in a variety of contexts to guide the development of policies, methods and standards for selecting and preserving authentic electronic records over time. The Strategies Task Force is designing the framework as a template that can be contextualized and used in a variety of national, multinational, international, cultural and organizational contexts.

Methodology

The Strategies Task Force utilizes two methods to conduct its research. First, derived from the findings and final recommendations of the Authenticity, Appraisal and Preservation Task Forces, the Strategies Task Force distills principles and criteria for the formulation of policies and standards. Second, based on the work of the other three Task Forces, the Strategies Task Force identifies

universal concepts, principles and methods that can be used in specific national, organizational and cultural contexts. The design, methodology and implementation of the research reflect a multidisciplinary, international and iterative collaboration among the Task Forces and international and national research teams.

Timeline and Action Plan

In February 2001, the Strategies Task Force met and designed a timeline and action plan to complete its research. The highlights of the timeline are included below:

October 2000

The Strategies Task Force convened. Developed research questions, goals and objectives, methodology and preliminary timeline.

February 2001

The Strategies Task Force developed action plan including timeline for submission of Task Force reports. Timeline presented to International Team, Task Forces, National and Multinational Teams. Research assistants identified international and national standards, policies and legislation.

June 2001

Task Forces produce first draft of reports including recommendations. International Team reviews, discusses and endorses the recommendations of the three Task Forces.

August 2001

The Strategies Task Force prepares an outline of the intellectual framework guiding the development of policies, strategies, procedures and standards as distilled from the recommendations of the other three Task Forces.

October 2001

Research Teams submit feedback papers. The Strategies Task Force articulates recommendations. National and multinational teams review recommendations and distillation of principles and criteria within their respective contexts and report the results to the International Team. The Task Forces finalize their reports to the International Team. The International Team endorses the draft distillation of the Strategies Task Force and receives the final reports of the research teams.

November-December 2001

The Strategies Task Force integrates the final reports of the Task Forces and national and multinational research teams and finalizes the intellectual framework guiding the development of policies, methods and standards.

Future Research

The research design of the Strategies Task Force builds on the existing framework, expertise and collaboration among InterPARES Task Forces, teams and researchers. In fact, the methodology and

focus of the Strategies Task Force provides for a synthesis and meta-analysis. The importance of the work of the Strategies Task Force can be measured in its contribution of an intellectual framework that can guide the development of policies, methods and standards for selecting and preserving authentic electronic records over time. Without this, the cultural and historical record and the value of today's information institutions including government and global electronic commerce may be lost. The distillation and identification of key principles from the research of the Authenticity, Appraisal and Preservation Task Forces is the first step toward the creation of an intellectual framework that can be used in a variety of international, national, organization and cultural contexts

Concluding Remarks

Ian E. Wilson
National Archivist of Canada

Ian Wilson has been Canada's National Archivist since July 1999. He was previously the Archivist of Ontario, where he was also responsible for the province's public library system. Prior to that, he served as the provincial archivist for Saskatchewan, Queen's University archivist and as archivist for the City of Kingston.

It is a privilege to be here today and to participate in this second InterPARES symposium, although I must admit that I feel like an interloper – all of those speaking today and many others in our audience have been working intensively all week on some very key issues that are of interest to all of us, they have been modeling and refining concepts and grappling with intangibles. You have been dealing with the arcane nature of CSIP (Case Study Interview Protocol) and TEDGI (Template Element Data Gathering Instrument), and only in discussions amongst archivists could references to "The Creator" refer to some low-level bureaucrat who filled in a form and signed his or her name.

I have been hearing a great deal about processes and the complex dynamics of international committee work. Today, we have heard about the outcome of these meetings and of the solid progress being

made in several areas. Most speakers have had the unenviable task of condensing five or eight days of discussion, thought and argument into twelve minutes. But, through all of this, I have a clear impression of the collaboration that is taking place and the advances that have been made.

I do not claim to understand all of what we have heard nor the full implications for record-keeping systems and for archives, but I do know that the investment of time, energy and professional experience made by my institution, by universities and by leading archives around the world is bearing fruit.

One point I was asked to deal with in these concluding remarks is the future of InterPARES research. That future will develop from two key strengths. Like the SSHRCC Review Committee which met here earlier this week, I have been struck by the extraordinary international collaboration that is taking place under the InterPARES umbrella. Dr. Duranti and her colleagues here have organized, motivated and inspired an unprecedented effort. Experts, representing different archival traditions from around the world, have been spending a great deal of time focussing on the most challenging problem facing archivists, lawyers, auditors and computer scientists – electronic records that maintain their integrity, reliability and authenticity over time. This is a timely collaborative effort searching for answers. I understand that the SSHRCC review panel members were impressed by how all presenters used the word “we” rather than “I" when

referring to this project. This is the fundamental strength of InterPARES and on it much may be built.

The other strength of this collaborative effort has been the ability of the project leaders to involve students here, at UCLA and elsewhere. We have met some of them today, especially those who presented their findings on aspects of the work of the Task Forces, an appropriate reflection, I think, of the active and important role they are playing on these Task Forces. I have met others in the last day or two. I would congratulate all of them for their participation and encourage them, wherever they may eventually work in the archival system, to continue their active interest in research.

Electronic records are not the only challenge that archives face; the electronic world has opened new possibilities for making the fragile multi-media documents found in archives - preserved, protected and too often hidden - available to a much wider audience. There is a fundamental transformation under way in archival services and we need fresh and innovative approaches to making the diverse and extensive holdings of Canadian archives available to our real clientele, the Canadian people from coast to coast to coast. But that is the subject of another talk.

Returning to the InterPARES initiative, staff at the National Archives have derived considerable benefit from InterPARES. Our participation has enabled us to test and refine our methodology and has suggested some new and innovative approaches. This is also an opportunity for me to recognize publicly the National Archives staff

who have been leading this initiative within our own institution: Bruce Walton, Yvette Hackett, Cathy Bailey and Normand Fortier. They, with a number of their colleagues who have worked on the case studies, have provided key support to InterPARES by contributing knowledge and experience about working directly with the very practical problems of electronic record-keeping systems.

Listening to today's speakers, it is clear that considerable progress has now been made in moving the InterPARES research agenda forward – we have a solid base of case study results to build on in elaborating the authenticity requirements which are to be one of the fundamental outcomes of the project. In the appraisal and preservation research domains, we have carefully-considered models of those archival activities, models which when finalized will document the methods by which the authenticity requirements can be incorporated into archival work. We are already seeing how some of these results can be converted into action:

1. the National Archives is currently examining the issue of what logical format to use for the transfer and storage of textual electronic records. One criterion against which candidate formats and conversion software will be measured is the authenticity of the converted records – how faithfully do they reproduce the essential characteristics of the source records? We are eagerly waiting for InterPARES to tell us what those essential characteristics are;

2. for appraisal purposes, we expect that the InterPARES findings will provide us with tools that we can use to evaluate more precisely the feasibility of preserving a given body of records. I should also note that this benefit to the National Archives will be a valuable addition to an already well-established appraisal *régime*; for other archives with less experience in this work, the output from the Appraisal Task Force will be even more beneficial; and

3. the National Archives is also keenly interested in the products of the Task Force which is looking at the preservation issue – we know that in a world of imperfect record-keeping practices and rapidly changing technology, there may never be a one-size-fits-all preservation solution; however, we are encouraged by the fact that the model of the preservation activity being developed identifies several different preservation options. In particular, we are looking at how those options, when combined with the authenticity requirements, will assist in the development of a preservation strategy that will guide us in our work with different types of records.

Looking ahead to the proposed second phase of InterPARES, we know that, even as we grapple with the preservation of "traditional" textual records in electronic form, the world of information technology is creating entirely new classes of digital materials which archives or others will be called upon in the future to preserve and to make available. The National Archives fully supports the second

phase of InterPARES. This support is based on our expectation that the considerable investment which the archival, academic and other communities have made in phase 1 will be fully justified by the results delivered later this year. It will be very important that the second phase do three things:

a) the first is to incorporate the methodological and organizational lessons learned from the current work and to implement them wherever we can;
b) that the second phase be a full continuation of the first – more than organizationally and methodologically – by incorporating into its thinking the lessons learned from the reaction to and comments on the results of the first phase, both positive and negative; and
c) finally, as with the current work, the second phase must aim not at the moving target of technology, but at the established target of archival principles as they are manifested in our understanding of what it means to preserve *records*.

As head of a major government archives, I must underline the reality that the second phase of InterPARES for the institutional partners must be implementation of phase 1. Others in government have referred to a crisis in government records management systems. Regardless of whether one agrees with this strong wording or not, there is clearly a sense of urgency developing around this issue. Throughout government, in Ottawa, in the provinces, in our larger municipalities as well as in the corporate world, it is clear that help is

needed on issues surrounding electronic records; that basic and fundamental rules of record-keeping have been forgotten or neglected. The issue, commonly referred to as Y2K, was essentially a failure of information management in dealing with electronic systems. Why did professional managers entrust their mission-critical program records or data to systems that would essentially expire at midnight on December 31, 1999? Other issues touching on the integrity of the official record appear on the front pages of newspapers with increasing regularity. The solutions offered by a legion of knowledge management consultants falter in many organizations as the basic information management infrastructure is found wanting or entirely absent. There are encouraging signs that some senior managers, at all levels of government, now realize that they face a serious challenge in program continuity and in public accountability.

The institutions our archives serve can and need to benefit from the research findings of the InterPARES Project. But to be of practical benefit, these results need to be translated into policies and procedures, systems and processes, and training packages that we can adopt and apply. The question in my mind is whether each institution will do this on its own or whether the collaborative approach, which has so distinguished the InterPARES Project to this point, will continue as we undertake the difficult implementation phase.