InterPARES Project

International Research on Permanent Authentic Records in Electronic Systems

Authenticity Task Force Final Report

Task Force Members:

Heather MacNeil, University of British Columbia (Chair)
Chen Wei, Beijing Municipal Archives
Luciana Duranti, University of British Columbia
Anne Gilliland-Swetland, University of California, Los Angeles
Maria Guercio, University of Urbino
Yvette Hackett, National Archives of Canada
Babak Hamidzadeh, University of British Columbia
Livia Iacovino, Monash University
Brent Lee, University of British Columbia
Sue McKemmish, Monash University
John Roeder, University of British Columbia
Seamus Ross, University of Glasgow
Wai-kwok Wan, Hong Kong Public Record Office
Zhao Zhon Xiu, State Archives of China

28 October 2001

ACKNOWLEDGEMENTS

The Task Force wishes to thank the InterPARES Project staff for its administrative support of the Task Force's work. Staff members are: Tahra Fung (University of British Columbia), Kevin Glick (University of Albany, State University of New York), Jean-Pascal Morghese (University of British Columbia), and Peter Van Garderen (University of British Columbia).

We would like to thank the research assistants at the Netherlands Institute for Archival Education and Research; State Archives of Rome; the University at Albany, State University of New York; University of British Columbia; University of California, Los Angeles; and the University of Glasgow. The research assistants are: Flora Anastassiou, Andrew Ashton, Chaja Beck, Tiran Behrouz, Lisa Beitel, Gijs Boon, Mirjam Brouwer, Vincenzo De Meo, Henk Duits, Robert Edwards, Anna Gibson, Prisca Giordani, Elaine Goh, Monica Grossi, Rebecca Hatcher, Robyn Hulley, Irene Kaplan, Ingmar Koch, Marta Maftei, Francesca Marini, Ian McAndrew, Shauna McRanor, April Miller, Susanna Oreffice, Eun Park, Marisol Ramos, Alex Richmond, Kalpana Shankar, Richard Sloma, Jacqueline Stroet, Melissa M Terras, Ciaran Trace, Silvia Trani, Claire Vesseirre, Marianne Vos, Robert van Vuuren, Lara Wilson, Joleen Wright, Jane Zhang.

We would also like to thank the following institutions, organizations, and their staff for their assistance in conducting the case studies: Academic Medical Centre, Utrecht; Azienda Municipale Ambiente, Rome; Banca d'Italia; Banca Nazionale del Lavoro; the Banff Centre for the Arts; Canadian Intellectual Property Office; Department of Indian Affairs and Northern Development, Government of Canada; Department of National Defence, Government of Canada; Department of the Solicitor General, Government of Canada; Lepera & Ward Architectural Firm; Ministry of Housing, Spatial Planning and the Environment, Government of the Netherlands; Ministry of Justice, Government of the Netherlands; the National Archives of Canada; Nanjing University; New York State Workers' Compensation Board; Pennsylvania State University, Province of North Holland, Haarlem; the Smithsonian Institution; United States Patent and Trademark Office; the United States National Archives and Records Administration; the University at Albany, State University of New York; the University of British Columbia; the University of California, Los Angeles; University of Glasgow.

We acknowledge gratefully the following agencies and institutions for their support: Banca d'Italia; Beijing Municipal Archives; the Central State Archives of Italy; the National Archives of Canada; the National Historical Publications and

Records Commission; the National Research Council of Italy; the Netherlands Institute for Archival Education and Research; the Smithsonian Institution; the Social Sciences and Humanities Research Council of Canada; the State Archives of Rome; the State Archives of China; the United States National Archives and Records Administration; the University at Albany, State University of New York; the University of British Columbia; the University of California, Los Angeles; the University of Glasgow.

INFORMATIVE ABSTRACT

The goal of the Authenticity Task Force was to identify conceptual requirements for assessing and maintaining the authenticity of electronic records. To achieve this goal, the Task Force adopted two distinct, yet related analytical approaches. The first approach was a theoretical and deductive one, based on contemporary archival diplomatics. The second approach was an inductive and empirical one that employed selected case studies of extant electronic systems. The primary outcome of the work of the Task Force has been the development of two sets of requirements: the first set includes requirements that support the presumption of the authenticity of electronic records before they are transferred to the preserver's custody; while the second set includes requirements that support the production of authentic copies of electronic records after they have been transferred to the preserver's custody.

The deductive and the inductive approaches employed by the Task Force have resulted in the construction of a detailed profile of the complexity of contemporary electronic records and in the identification of the extent to which the records are embedded within the specific juridical-administrative, provenancial, procedural, documentary, and technological contexts in which they are created. The Task Force found that most contemporary records systems are a hybrid of electronic and paper records; that few explicit measures are employed to ensure the authenticity of electronic records, and that authenticity is generally assured through procedural means. While it was successful in developing a conceptual framework for establishing the requirements for preserving authentic electronic records, the Task Force did not succeed in creating a single, comprehensive typology of authenticity requirements for electronic records. It did, however, identify possible perspectives from which a typology could be constructed and which merit further exploration. In the view of the Task Force, a typology based upon individual creators and the acts/procedures/functions they carry out is likely to be the most effective starting point in any typification of electronic records.

The Task Force also found that the complexity of electronic records and record-keeping made it extremely difficult for researchers to identify a single, appropriate unit of analysis. While the two analytical approaches adopted by the Task Force in carrying out its research contributed to an understanding of the nature of the record and its long-term preservation, an overall systems approach, one that takes into account the total record-keeping environment, is also needed.

1. INTRODUCTION

This report communicates the results of the work of the InterPARES Authenticity Task Force. The charge of the Authenticity Task Force was to identify conceptual requirements for assessing and maintaining the authenticity of electronic records. The original InterPARES research plan identified five questions that were to be addressed within the Authenticity Domain (Domain 1):

- What are the elements that all electronic records share?
- What are the elements that allow us to differentiate between different types of electronic records?
- Of those elements, which will permit us to verify their authenticity over time?
- Are the elements for verifying authenticity over time the same as those that permit us to verify their authenticity in time, that is, at the point at which they are originally created and transmitted?
- Can the elements be removed from where they are currently found to a place where they can more easily be preserved and still maintain the same validity?

As this report describes, however, these initial questions were considerably revised and refined, and new, unanticipated questions emerged, during the course of the research.

Recognizing the need to delineate the full complexity of the issues associated with the authenticity of electronic records, the Task Force sought to triangulate two distinct yet complementary research approaches. The first approach was a theoretical and deductive one, based on contemporary archival diplomatics. It involved identifying and defining the elements of an ideal electronic record in general, and those that are relevant to a consideration of its authenticity in particular, using concepts and methods derived from diplomatics and archival science that in turn are based upon what is known about traditional records, juridical systems, and record-keeping practices. The second approach was an inductive and empirical one that employed selected case studies of extant electronic systems. While these systems were in many cases far removed from the ideal electronic record as established through the first approach, they were able to elucidate the shifting boundaries of electronic records, emergent record-keeping processes, and new manifestations of traditional record elements. Both approaches were aimed at theory building, and the conceptual requirements for assessing the authenticity of electronic records emerged out of their triangulation.

The primary outcome of the work of the Authenticity Task Force has been the development of two sets of requirements: the first set includes requirements that support the presumption of the authenticity of electronic records *before* they are transferred to the preserver's custody; while the second set includes

requirements that support the production of authentic copies of electronic records *after* they have been transferred to the preserver's custody. The research also resulted in several additional datasets and products, and these are also discussed in this report.

2. BASIC PREMISES OF THE RESEARCH

2.1 Definition of record

A record is defined as any document made or received and set aside in the course of a practical activity. The interpretation of this definition in the context of electronic systems is discussed in sections three and four of this report.

2.2 Definitions of authenticity, authentic, and authentic record

In common usage, the concept of *authenticity* is defined as "the quality of being authentic, or entitled to acceptance,"[1] while the term a*uthentic* means "worthy of acceptance or belief as conforming to or based on fact" and is synonymous with the terms genuine and bona fide. *Genuine* "implies actual character not counterfeited, imitated, or adulterated [and] connotes definite origin from a source." *Bona fide* "implies good faith and sincerity of intention."[2] From these definitions it follows that an *authentic record* is a record that is what it purports to be and is free from tampering or corruption.

2.3 Rationale for establishing conceptual requirements for assessing the authenticity of electronic records

In both archival theory and jurisprudence, records upon which the creator relies in the usual and ordinary course of business are presumed authentic. However, records created and maintained in electronic form are continually at significant risk of inadvertent or intentional alteration, and such alteration may also not be readily perceptible. The authenticity of electronic records is threatened whenever they are transmitted across *space* (that is, when sent between persons, systems or applications) or *time* (that is, either when they are stored offline, or when the hardware or software used to process, communicate, or maintain them is upgraded or replaced). Requirements for assessing the authenticity of electronic records that are preserved over the long term are necessary, therefore, to support the presumption that an electronic record is, in fact, and continues to be, what it purports to be and has not been modified or corrupted in essential respects. The interpretation of what constitutes "in essential respects" is explained in section four of this report.

[1] *Oxford English Dictionary,* 2nd ed., s.v. "authenticity".
[2] *Merriam-Webster Online Dictionary*, s.v. "authentic".

2.4 Differentiating between authenticity and authentication

Because of the ongoing developments in the area of authentication technologies, it is necessary to clarify the distinction between *authentication* and *authenticity*, which is the focus of InterPARES. In common usage, *authentication* is understood as a declaration of a record's authenticity at a specific point in time by a juridical person entrusted with the authority to make such declaration. It takes the form of an authoritative statement (which may be in the form of words or symbols) that is added to or inserted in the record attesting that the record is authentic.

Digital signature and public key infrastructure (PKI) are examples of technologies that have been developed and implemented as a means of authentication for electronic records that are transmitted across space. Although record-keepers and information technology personnel place their trust in authentication technologies to ensure the authenticity of records, these technologies were never intended to be, and are not currently viable as a means of ensuring the authenticity of electronic records over time.

3. RESEARCH DESIGN AND METHODOLOGY

3.1 The theoretical-deductive approach

In the first stage of its research, the Task Force established the theoretical framework for the analysis of various types of electronic records and the identification of those elements that need to be preserved to ensure the records' authenticity over time. The *Template for Analysis* embodies this framework. The *Template* is a decomposition of an electronic record into its constituent elements.[3] The decomposition defines each element, explains its purpose, and indicates whether, and to what extent, that element is instrumental in assessing the record's authenticity.

The theoretical perspective that shaped the development of the *Template* was contemporary archival diplomatics. Diplomatics emerged in the seventeenth

[3] The *Template for Analysis* is available on the InterPARES website at http://www.interpares.org/reports.htm. The term "elements" is used differently in diplomatics to the way in which it is used in information systems design. In developing the initial research questions and the *Template for Analysis*, the Task Force used the diplomatic term "elements" to refer to both general and specific characteristics of a record that may be found in its documentary form, in annotations, or in one or more of its various contexts. As the research progressed, however, the Task Force found it necessary to narrow the scope of the concept. In the *Requirements for Authenticity*, therefore, the term 'record elements" refers specifically to the intrinsic and extrinsic elements of a record's documentary form as these are identified in the *Template for Analysis*. Such redefinition is illustrative of how diplomatics continues to evolve in response to the changing nature of the record.

century as an analytical technique for determining the authenticity of records issued by sovereign authorities in previous centuries. Its primary purpose was to ascertain "the reality of the rights or truthfulness of the facts"[4] contained in such documents. The tenets and methods of diplomatics were laid out in 1681 in a treatise written by a Benedictine monk, Jean Mabillon. Mabillon examined, among other things, the language of the documents, their characteristic parts, their seals, and the systems of chronology used in dating them. On the basis of this examination, "Mabillon stated what, for a particular time and place, was the correct form for a genuine document, and presented the general principles of diplomatics."[5] The original use of diplomatics was to determine a record's authenticity for legal purposes and that use continued into the eighteenth century when its concepts and principles were incorporated into the curriculum of many European faculties of law. By the end of the nineteenth century, however, under the influence of classical philology and the scientific school of historiography, diplomatics emerged as a tool for assessing the authority of medieval records as historical sources.

Over the last twenty years there have been numerous calls from within the archival community to revive and adapt diplomatics as an aid to understanding the record-keeping processes of contemporary bureaucracies. Delegates to the 1989 International Council on Archives' Second European Conference on Archives, for example, recommended "that the development of the discipline of modern diplomatics be promoted through research in the typology of contemporary records and in the records-creating procedures of contemporary institutions." [6] In Europe, notable archival efforts to construct a modern diplomatics include the work undertaken by Dutch archivists to develop a typology of records created by organizations since the nineteenth century in the Netherlands;[7] and the adaptation of traditional diplomatic concepts and methods to the record-keeping environment of contemporary Italian administration undertaken by Paola Carucci.[8]

---

[4] Luciana Duranti, "Diplomatics: New Uses for An Old Science," *Archivaria* 28 (Summer 1989): 17.

[5] James Westfall Thompson, *A History of Historical Writing* (New York: The Macmillan Co., 1942), vol. 2, 19.

[6] Judith Koucky, ed. "*Second European Conference on Archives: Proceedings* (Paris, International Council on Archives, 1989), 113. The delegates' recommendation was in support of comments made by Francis Blouin. See Francis X. Blouin, Jr., "Convergences and Divergences in Archival Tradition: A North American Perspective," *Second European Conference on Archives*, 28-29. Other archivists who have advocated the revival of diplomatics for modern records include Tom Nesmith, "Archives from the Bottom Up: Social History and Archival Scholarship," *Archivaria* 14 (Summer 1982): 5-26; Don C. Skemer, "Diplomatics and Archives," *American Archivist* 52 (Summer 1989): 376-82; and Hugh Taylor, "My Very Act and Deed: Some Reflections on the Role of Textual Records in the Conduct of Affairs," *American Archivist* 51 (Fall 1988): 456-69.

[7] For a summary of this research and its products, see David Bearman and Peter Sigmond, "Explorations of Form of Material Authority Files by Dutch Archivists," *American Archivist* 50 (Spring 1987): 249-53; Peter J. Sigmond, "Form, Function and Archival Value," *Archivaria* 33 (Winter 1991-92): 141-47.

[8] Paola Carucci, *Il Documento Contemporaneo* (Rome: La Nuova Italia Scientifica, 1987).

In North America, the most comprehensive effort to adapt traditional diplomatics to contemporary record-keeping practices is embodied in the work of Luciana Duranti of the University of British Columbia. In a series of articles written between 1989 and 1992,[9] Duranti examined the principles and concepts developed by diplomatic theorists for evaluating the authenticity of medieval documents to determine whether they could be adapted for application to the records generated by modern bureaucracies. Over the course of the six articles, she refined and reinterpreted the classical concepts, and introduced new ones to take into account the variety and complexity of bureaucratic record-keeping environments.

Duranti's series of articles resulted in a preliminary elaboration of contemporary archival diplomatics, an adaptation of traditional diplomatic concepts and methods to contemporary record-keeping environments, and an integration of these concepts and methods with those of archival science. It also laid the groundwork for a research project carried out between 1994 and 1997 at the University of British Columbia entitled *The Preservation of the Integrity of Electronic Records* ("the UBC project").[10] The goal of that project was to identify and define conceptually the nature of an electronic record and the conditions necessary to ensure its integrity (that is, its reliability and authenticity) during its active and semi-active life. The research resulted in a set of standards and rules for developing and implementing a trustworthy electronic record-keeping system.[11]

The elements of an electronic record identified in the UBC Project provided the starting point for the identification of the InterPARES *Template* elements. Based on researcher input from a range of disciplinary perspectives, as well as data collected during the case studies, these original elements were revised and extended, and new elements were added as the research progressed. For example, the broader administrative and documentary contexts in which a record is created, handled, and maintained were more precisely articulated in the *Template* than they had been in the UBC Project, and a new category of context, that is, technological context, was identified and elaborated.

---

[9] Duranti. "Diplomatics I," 7-27; "Diplomatics ... (Part II)," *Archivaria* 29 (Winter 1989-90): 4-17; " Diplomatics ... (Part III)," *Archivaria* 30 (Summer 1990): 4-20; "Diplomatics ... (Part IV)," Archivaria 31 (Winter 1990-91): 10-25; "Diplomatics ... (Part V)," *Archivaria* 32 (Summer 1991): 6-24; "Diplomatics ... (Part VI)," *Archivaria* 33 (Winter 1991-92): 6-24. Published in a single volume as Duranti, *Diplomatics: New Uses for an Old Science* (Lanham, Maryland, and London: Scarecrow Press in association with the Society of American Archivists and Association of Canadian Archivists, 1998).

[10] For an overview of the findings of the UBC Project see Luciana Duranti and Heather MacNeil, "The Protection of the Integrity of Electronic Records: An Overview of the UBC-MAS Research Project," *Archivaria 42* (Fall 1996): 46-67.

[11] The outcomes of the UBC Project were subsequently substantially incorporated into the Design Criteria Standard for Electronic Records Management Software Applications (DOD 5015.2-STD) promulgated by the U.S. Department of Defense.

To assist the researchers' understanding of traditional diplomatic elements and their contemporary interpretation, student research assistants traced the lineage of the elements included in the *Template* back to their original elaboration in the work of traditional French, German, and Italian diplomatists.[12] The researchers reasoned that a sound understanding of the historical meaning of the elements would better equip them to assess their contemporary relevance. Research assistants also prepared a sample typology of papal chancery documents to facilitate the researchers' understanding of how traditional diplomatists viewed the relationship between authenticity and documentary form, and, more specifically, how individual elements of documentary form supported the attestation of a record's authenticity.[13]

Viewed from the perspective of contemporary archival diplomatics, 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,[14] 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 (meaning its creation may be mandatory or discretionary), and at least three persons (author, writer, and addressee) are involved in its creation (although these three conceptual persons may in fact be only one physical or juridical person).

In a traditional record-keeping environment, these characteristics manifest themselves in explicit and implicit ways. For example, the name of the author may appear on the letterhead; and the archival bond may be expressed in a classification code or some other unique identifier that appears on the face of a record. The purpose served by these individual elements also depends on their specific form of expression. For example, the identification of the name of the author that appears in the letterhead serves the purpose of identifying aspects of the record's provenancial context. When that same name appears as a signature at the bottom of the record, it serves the purpose of attesting to the validity of the record or its content. **The working hypothesis of the Authenticity Task Force was that, while they may manifest themselves in different ways, these same or similar elements are present, either explicitly or implicitly in electronic records**. The *Template for Analysis* was created to test that hypothesis. The elements of an electronic record included in the

[12] The document showing the lineage of elements included in the *Template for Analysis* is available on the InterPARES website.
[13] The sample typology is available on the InterPARES website.
[14] According to the Authenticity Task Force's *Research Methodology Statement*, a fixed form "means that (1) the binary content of the record, including indicators of its documentary form, are stored in a manner that ensures it remains complete and unaltered; and (2) technology has been maintained and procedures defined and enforced to ensure that the content is presented or rendered with the same documentary form it had when it was set aside." The Statement is available on the project website.

Template fall into four main categories: *documentary form*, *annotations*, *context*, and *medium*:[15]

Documentary form is defined as the rules of representation according to which the content of a record, its immediate administrative and documentary context, and its authority are communicated. It 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:
  1) elements that convey aspects of the record's juridical and administrative context (for example, the name of the author, addressee, the date);
  2) elements that communicate the action itself (for example, the indication and description of the action or matter);
  3) elements that convey aspects of the record's documentary context and its means of validation (for example, 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 (for example, textual, graphic, image, sound, or some combination of these);
  - specific presentation features (for example, special layouts, hyperlinks, colors, sample rate of sound files);
  - electronic signatures and electronic seals (for example, digital signatures);
  - digital time stamps;
  - other special signs (for example, digital watermarks, an organization's crest or personal logo).

Annotations (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:

1) additions made to the record after its creation as part of its execution, for example, the date and time of transmission added to an email record at the moment it is sent, or the indication of attachments added before it is transmitted;

[15] For a more detailed discussion of the *Template for Analysis*, see Heather MacNeil, "Providing Grounds for Trust: Developing Conceptual Requirements for the Long-Term Preservation of Authentic Electronic Records," *Archivaria* 50 (Fall 2000): 56-67.

2) additions made to the record in the course of handling the business matter in which the record participates, for example, comments noted on the face of the record, or embedded in it, and dates of transmission to other offices;
3) 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 forth.

Context shifts the analysis away from the record itself to the broader structural, procedural, and documentary framework in which the record is created and managed. The identified elements of context correspond to a hierarchy of frameworks ranging from the general to the specific. They include the record's *juridical-administrative context*, its *provenancial context*, its *procedural context*, its *documentary context*, and its *technological context*. Knowledge of these elements is critical to an understanding of 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.[16]

Medium proved to be a problematic construct from the perspective of diplomatic analysis. In identifying and positioning the elements included in the *Template for Analysis*, the Authenticity Task Force struggled with the question of whether to treat the *medium*, that is, the physical carrier on which a record is stored, as a part of the record itself or as part of its technological context. For diplomatists examining medieval documents, the medium is an essential component of a record because the examination of the physical carrier on which the document is inscribed is one of the most obvious proofs of its authenticity.[17] In the translation of traditional diplomatic concepts into modern paper-based record-keeping environments, the medium has continued to be treated as a part of the record itself, mainly because the medium and the message are inextricably linked. The question was whether, in an electronic record-keeping environment, the medium should continue to be treated as an essential part of the record itself given that: (1) the medium and the message are no longer inextricably linked; (2) what is inscribed on or affixed to the medium is not a record as such (or words, or pictures), but a bitstream; and (3) the choice of a medium by those creating or maintaining the record is often arbitrary and carries no particular significance.

[16] For a discussion of the embeddedness of electronic records within these contexts, see Anne J. Gilliland-Swetland and Philip Eppard, "Preserving the Authenticity of Contingent Digital Objects: The InterPARES Project," *Dlib Magazine*, 6 July/August 2000. Available at: <http://www.dlib.org/dlib/july00/eppard/07eppard.html>.

[17] For example, a royal diploma of Childebert I (King of Francs, 6th century) that is written on parchment instead of papyrus is considered false. The medium also provides evidence of the manner in which medieval documents were prepared. The documents from the German chancery have many erasures and corrections in comparison to the documents of the papal chancery, indicating a lesser degree of care and accuracy in the preparation of the final documents.

It is taken for granted that a record is a representation of a fact or act that is memorialized on a physical carrier, that is, a medium, and preserved by a physical or juridical person in the course of carrying out its activities.[18] 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, that is, the source of the record, cannot endure for any length of time unless it is affixed to a medium. Storage of a bitstream on a disk or tape, however, while necessary for the bitstream to endure, is not sufficient to preserve a record as a record. As the Preservation Task Force observed early on in its deliberations, "strictly speaking, it is not possible to preserve an electronic record. It is only possible to preserve the ability to reproduce an electronic record. It is always necessary to retrieve from storage the binary digits that make up the record and process them through some software for delivery or presentation."[19] Moreover, while affixing a bitstream to a medium is a pre-condition to the existence of an electronic record, this does not mean that it is a relevant factor in assessing 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. By the end of its research, therefore, the Authenticity Task Force concluded that the medium should be considered part of the record's technological context, rather than an essential part of the record itself.

Initial development of the *Template* took place over a nine-month period from January to September 1999. During that time, the *Template* was revised numerous times by both the Authenticity Task Force and the InterPARES International Team. By June 1999, the *Template* was considered sufficiently developed to begin the process of testing and refining it through case studies of real-life electronic systems.

3.2 Empirical-Inductive Approach

As discussed above, the *Template for Analysis* began as a model of an ideal record that, based upon prior archival knowledge of record types, delineated all the possible known elements that a record may contain. However, where diplomatic typologies and analysis have in the past been developed retrospectively based upon what was known about existing records, one goal of InterPARES was to develop a predictive model that would assist archivists in identifying future record types and the necessary requirements for maintaining their authenticity over time. In the first year of the project, InterPARES researchers determined that they could develop a richer picture of the complex nature of electronic records if they triangulated the theoretical, deductive

[18] 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.
[19] 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.

diplomatics-based approach with an inductive, empirical approach that was based on an examination of actual electronic records and electronic record-keeping systems. This examination was conducted by means of purposively selected, interpretive case studies of electronic systems that contained, or were deemed likely to contain, electronic records. These case studies were directed towards understanding electronic records within their various contexts as well as the relationships of those contexts to each other.

While the addition of this "bottom-up" approach extended InterPARES research activities considerably beyond those originally envisaged, it provided a rich dataset that informed the theoretical development by indicating the increasing role of procedural and technological context in ensuring and maintaining the authenticity of records. At the same time, the application of the *Template of Analysis* to existing records and record-keeping systems was able to indicate which necessary extrinsic and intrinsic elements of form were not present in systems as they were currently designed and operating, thus demonstrating potential weaknesses or deficiencies in the records or record-keeping systems examined.

3.2.1 Use and selection of case studies

The Task Force researchers adopted a grounded theory approach in which four successive rounds of case studies of electronic systems that contained or potentially contained records were examined in order to identify and describe phenomena associated with the records and their contexts. Grounded theory is a method for discovering concepts and hypotheses and developing theory directly from data under observation.[20] Cases are selected for study "according to their potential for helping to expand on or refine the concepts or theory that have already been developed. Data collection and analysis proceed together."[21]

Because of the grounded theory approach, researchers employed theoretical, rather than statistical sampling in the selection of case studies. Glaser and Strauss describe the process of theoretical sampling as "a process of data collection for generating theory whereby the analyst jointly collects, codes, and analyzes his data and decides what data to collect next and where to find them, in order to develop his theory as it emerges."[22] In other words, Task Force researchers purposively identified the cases that seemed most likely to elucidate phenomena that the research was seeking to understand (for example, what happens to active or inactive electronic records when they are subject to migration?). No attempt was made to draw a representative or statistically significant sample. In the first two rounds of case studies, the case studies

[20] Barney G. Glaser and Anselm L. Strauss, *The Discovery of Grounded Theory: Strategies for Qualitative Research* (Chicago: Aldine Atherton, 1967), 6-7, 46.
[21] Steven J. Taylor and Robert Bogden, *Introduction to Qualitative Research methods: The Search for Meanings*, 2nd ed. (New York: Wiley, 1984), 126.
[22] Glaser and Strauss, *Grounded Theory*, 45.

focused on *electronic* systems, although a considerable amount of contextual data was collected to elucidate the broader *record-keeping* environment. Following the International team and Authenticity Task Force's evaluation of these case studies and analysis of case study data, the criteria for selection were adjusted to support continued theory building.

The data gathered through these case studies was then used to test and extend the *Template for Analysis*. The translation of the case study data into a form that could be analyzed diplomatically by the *Template* was achieved by coding the data for inter-related themes and concepts using a *Template Element Data Gathering Instrument (TEDGI)*. The data collected through the case studies was also made available to the Appraisal and Preservation Task Forces to assist them with modeling preservation processes and then walking through their models.

First and second round case studies had to meet at least three of the following criteria:

1) Systems that contain, generate, or have the potential or possibility of generating records.[23]
2) Systems that have gone through one or more migrations.
3) Systems where migration(s) was (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 the pre-migration and the post-migration versions were available and functional.
6) Systems for which detailed documentation (design, implementation, migration, metadata) exists.
7) Systems with a diversity of information configurations (for example, contain both text and images).

In addition to these selection criteria, among the candidate systems proposed by the same archival institution, an effort was made to ensure diversity in content and type of records (that is, case studies representing a variety of systems proposed by the same institution). Between institutions, an effort was made to identify and conduct case studies on record-keeping systems performing similar functions (for example, student registration systems in different universities). The researchers believed that both of these factors might enable them to see

[23] As explained in section 3.1, p. 6, above, a record 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.

emergent patterns relating to the nature of organizational record-keeping and specific record-keeping functions.

A key issue encountered by the researchers, and indeed by any archivist or records managers who works with electronic records, is the difficulty in identifying actual electronic records and their parameters. This issue stems from the nature of digital information systems, which are frequently multi-purpose, highly networked database systems that can contain a diversity of information elements that can be compiled and presented in a variety of ways (for example, through hardcoded report formats, stylesheets, and virtual "on-the-fly" views) and which can invoke a range of functionalities, according to the needs of different users. A single system may contain only raw data or information, one or more than one types of record, or a combination of record types and data or information. The diplomatic analysis of first and second round case studies indicated that few of the systems appeared to contain records that came close to the ideal promulgated in the *Template* (some systems proved to be information systems not containing records at all, while some contained records that were able to achieve their purpose but were not intrinsically very good records). In line with the grounded theory approach, based upon what they had learned from the first *two* rounds of case studies, the researchers modified the case study selection criteria for the third and fourth round of case studies, to define more precisely the types of cases in which they were now interested. Through this redefinition, the following indicators of systems that are known to create records or have the potential to create records were incorporated:

- if the action in which the system participates is juridically required;
- if there is a business procedure in place to carry out that action;
- If the system operates within the management or strategic decision-making levels of the organization.

For case study rounds two to four, the researchers decided to examine only live systems (that is, systems still being actively used by the creator to carry out business activities), since the case studies of inactive electronic records indicated that too much contextual information had already disappeared for the Task Force to be able to analyze the records and record-keeping systems successfully. The researchers also eliminated criteria that related to systems and records that had undergone migration, since these had not proven to yield significant additional insights for either the Authenticity Task Force or the Preservation and Appraisal Task Forces. Additional desirable criteria identified for rounds 3 and 4 case studies were that:

1. Systems that come from different hierarchical levels within an organization; and optimally, systems supporting management and strategic level activities
2. Systems that contain supporting and narrative records
3. Systems from the private sector

4. Financial management systems
5. Multimedia systems
6. Computer-aided design (CAD) systems

3.2.2 Case study data

Between Spring 1999 and Spring 2001, four rounds of case studies were conducted by institutional and student researchers in government, university, and corporate agencies in Canada, the United States, Italy, the United Kingdom, the Netherlands, and China. The case studies included large-scale databases (such as patent and student registration systems), geographic information systems, and interactive web-based applications as they existed at the time when the case studies were conducted.

A drawback of any research that employs multiple selective case studies is the limited degree to which it is possible to compare across or generalize from individual case studies. Each case is highly sensitive to its own national, juridical, institutional, and technological contexts. Moreover, InterPARES case studies were conducted under a range of different conditions by different investigators. As a result, each case study had to be selected and analyzed on its own merits for how it might inform theory development by the researchers, and it was necessary to be cautious about the extent to which one could look for patterns emerging across case studies in similar institutional settings or performing similar functions in different settings. In an effort to control as much as possible the individual differences between case studies and within case study rounds, a *Case Study Interview Protocol (CSIP)* was developed by the Authenticity Task Force to standardize the interview process for the case studies as well as to provide data for populating the *TEDGI.* Several project investigators who would be conducting the case study interviews also participated in training sessions at UBC or UCLA on how to conduct the case studies as well as how to complete the *TEDGI* and *CSIP*.[24]

The *CSIP* (essentially the interview script) was divided into five sections: Context (juridical-administrative, provenancial, procedural, and documentary), Intrinsic Elements of Form, Extrinsic Elements of Form, Annotations, and Medium and Technological Context. A range of standardized questions was asked to elucidate each aspect. The same question was sometimes asked in different ways within the same section to check for consistency in responses. The same question was also sometimes asked in a different way in more than one section to identify any alternate perspectives of respondents with different backgrounds (for example, records managers and systems personnel). Interviewers, predominantly institutional archivists or archival science students participating in

[24] As required by the different researchers' individual institutions, the entire protocol for the case studies, and all subsequent revisions to the protocol was submitted for review and approved by the institutional review boards/offices for the protection of human subjects. The CSIP and TEDGI are available on the InterPARES website.

InterPARES, sought out respondents who were the records creators, records managers, and systems personnel primarily responsible for working with the electronic systems under study. Due to local requirements and practicalities, some interviews were with individuals, and some with groups of individuals. In some case studies, multiple interviews with different individuals were held. Interviewers also collected supporting documentation such as technical documentation, organization charts, and workflow rules; and sometimes followed up with interviewees when further information was required. The interviewers were then responsible for translating the data they had gathered through the CSIP and supporting documentation into the TEDGI, and for transmitting copies of all the case study data to both UBC and UCLA for analysis. In the third round of case studies, researchers at UBC were responsible for compiling the TEDGI.

Version 2.1 of the CSIP and 1.0 of the TEDGI were used for the first round of case studies. After the first round of case studies, the CSIP, TEDGI and the *Template for Analysis* were revised to eliminate, clarify, or expand aspects identified as problematic in the first round of case studies. In the second round of case studies, however, most interviewers still used version 1.0 (rather than 1.1) of the TEDGI, but version 3.0 of the CSIP. In the third round of case studies, researchers used version 1.1 of the TEDGI, and version 3.1 of CSIP. In the fourth round, researchers used version 1.1 of the TEDGI and version 3.2 of the CSIP. In total, data were analyzed for twenty-six completed case studies from the four rounds of case studies using two different versions of the TEDGI and four versions of the CSIP.

Multiple types of data were sought or created by the Task Force researchers in the course of conducting and analyzing each case study. These types included the CSIP and TEDGI, audio and videotapes of interviews, supporting procedural and technological documentation, and case study overviews. Not all data types exist for each case, however, due to variations in how data were collected (for example, interviewees could decline to be audiotaped), or to lack of availability of specific supporting technological or procedural documentation or translations of that documentation into English. It is also important to note that in the majority of cases, although the case study focussed on the electronic system, the actual record-keeping system comprised both paper and electronic components.

3.2.3 Case study data analysis

Each round of case studies was described and analyzed from the perspective of contemporary archival diplomatics—the primary emphasis of the work of the Authenticity Task Force—as well as through the application of analytical methods drawn from the social sciences. The rationale behind subjecting case study data to such a barrage of analyses was to render the most complete picture possible of the complexities of the modern electronic record, and to feed this emerging knowledge into the development of records theory, and archival diplomatics in particular.

3.2.3.1 Diplomatic analysis of case studies

The primary purpose of analyzing the case studies from the perspective of contemporary archival diplomatics was to consolidate information from case study documentation that would be relevant to, and required for, the drafting of the conceptual requirements for assessing the authenticity of electronic records, as well as for the development of a typology of electronic records based on those requirements. The analyses were undertaken by student researchers in their second year of the Master of Archival Studies Program at UBC. All the students were familiar with diplomatic analysis, having completed a course in diplomatics during their first year of the program.

The process of analysis took place in three phases: case studies from rounds one and two were analyzed in the first phase (Fall 2000); those from round three were analyzed in the second phase (January 2001); and those from round four were analyzed in the third phase (May 2001).[25] The process consisted of periodic team meetings of the research assistants with the Authenticity Task Force representative (Luciana Duranti) and the Project Coordinator (Tahra Fung) to discuss findings and brainstorm; and independent work by pairs of research assistants in the interim periods between team meetings. The work process differed from phase to phase as research assistants began to work more independently and as the responsibilities assigned to them grew. For example, responsibility for populating TEDGIs, which was assigned to case study researchers in rounds one and two, was assigned to the research assistants for rounds three and four case studies. The process produced a considerable amount of documentation, including synopses of the systems, answers to assigned questions, inquiries directed to case study researchers, and speculative scenarios. The final product of the diplomatic analysis was constituted by a "Final Report" for each system analyzed. Final reports were written for twenty-two of the twenty-six completed case studies. Four case studies were excluded from the diplomatic analysis due to insufficient information or language difficulties with technical documentation.

The diplomatic analysis of the first two rounds of case studies commenced in September 2000. Each pair of research assistants was assigned a case study and asked to complete a diplomatic analysis of the electronic system on the basis of the *Template for Analysis*; the *CSIP*, case study supporting documentation (including organization charts, lists of employee responsibilities, print-outs of screen views, interview tapes, legislation, diagrams of business procedures, glossaries used by the organization, etc.); and Duranti's six-part exploration of diplomatics.[26] After a preliminary examination revealed significant inconsistencies and differing interpretations on the part of interviewers in the

[25] The account of the process of diplomatic analysis is based on a summary prepared by Ian McAndrew, with contributions from April Miller and Anna Gibson.
[26] See above, fn. 9.

translation of data from the *CSIP* into the *TEDGI,* it was decided that only the *CSIP*s and not the *TEDGI*s completed by the case study researchers would be used in the diplomatic analysis in the first phase.

To complete the diplomatic analysis, the research assistants were assigned the following questions:

- How many records are in the system?
- What is the function of these records (that is, dispositive, probative, supporting, narrative[27])?
- What is/are the action(s) associated with the system?
- What types of documentary forms are included?
- What is the status of transmission of each documentary form (that is, original, draft, copy[28])?

Considering these questions was essentially a process of examination that led to four more questions:

- Does the system contain records?
- Should the system contain records?
- Is the system itself a record ?
- With the nature and function of the system in mind, is there a presumption of authenticity? If yes, what is the basis for this presumption?

Answering these questions proved considerably more challenging than had been anticipated. The fundamental problem the research assistants faced was that of identifying an electronic record in diplomatic terms. Although the research assistants had experience with the process of diplomatic analysis, they had only ever dealt with traditional paper records. To analyze the case studies in diplomatic terms, it was necessary first to penetrate the complexity of the electronic system and the surrounding record-keeping environment in order to establish whether records even resided within that system and, if so, to understand the specific ways in which they manifested themselves. To reach that understanding required a detailed knowledge of the electronic system and the record-keeping environment that was difficult to achieve. The difficulty stemmed in part from the fact that the knowledge had to be gleaned, not on the basis of an examination of the system itself and the entities within it, which is the traditional diplomatic approach; but, rather on the basis of the information found in the case study tools and related documentation.

While documentation from all sources was valuable in supporting the analysis of a given case study, the *CSIP* and its supporting documentation did not provide enough information to enable the research assistants to gain a good understanding of the relationship between the electronic system and the

[27] For the definition of these terms, see below, sec. 3.2.3.2, pp. 20-21.
[28] For the definition of these terms, see below, fn. 40.

business processes associated with it, and the relationship between the business processes and the types of records generated from them. Moreover, the supporting documentation was only included at the discretion of each interviewer conducting the case studies, who encountered issues not unfamiliar to institutional archivists responsible for appraising electronic records—up-to-date technological or procedural documentation may not exist, organizations or their units may be reluctant to provide copies of systems documentation for security or other reasons, and existing documentation may be intellectually inaccessible to the archivist for technical or language reasons. For some case studies, therefore, there was considerable supporting documentation, for others there was little or none. More supporting documentation did not always imply more and better information about the systems, however: some supporting documentation was very hard to understand and, in some cases, it was not clear why it was included at all. To fill in the gaps in their knowledge, the research assistants solicited the assistance of the interviewers who conducted the case studies. In some cases the interviewers were able to answer their questions; in others, however, they were either unable to obtain the needed information, or unable to obtain it within the time frame necessary.

As the analysis proceeded, it became increasingly clear that most of the systems under examination did not contain records, or at least, did not contain “good” records, when measured against the criteria established by contemporary archival diplomatics. In most cases this was because the entities identified within the electronic system did not appear to possess either a fixed documentary form or a stable content. To probe this situation further, UBC researchers decided to draft “scenarios” for certain cases. For those case study systems that had been found on first analysis to contain records (11, 15, and 19), research assistants were instructed to answer the questions already devised. However, since applying these questions to systems without records would not be practical, three contingency formats were designed. First, for case studies 06 and 10, a brief report was drafted introducing the system and explaining that no diplomatic analysis could be performed. Second, reports for case studies 04, 07, and 08 were drafted on the basis of a scenario positing that “the system does not contain records, but if it did they could be analyzed diplomatically as follows … ” Third, reports for case studies 01, 02, and 12 were drafted on the basis of a scenario positing that “the system does not contain records, but it should; it could be reconfigured such that it would contain records, as follows, and if this were done, the records could be analyzed diplomatically as follows ...”

Given that the case studies so far had yielded very little information useful for the formulation of the requirements for assessing authenticity, the researchers also decided to incorporate into the analysis the procedural rules for creating and maintaining reliable and authentic electronic records that had been developed by the UBC project. The data gathered from each case study concerning the methods used by the creator to support its presumption of the authenticity of the records in the system under examination were compared with the procedural

rules for creating and maintaining authentic records laid out in the UBC project.[29] On the basis of this comparison, the research assistants described the means currently in place that, from the creator's perspective, supported a presumption of record authenticity and identified additional methods for supporting and strengthening such presumption, based on the procedural rules.

The development of hypothetical case studies and the comparative analysis of real world data with the UBC procedural rules enabled the team to draft a preliminary set of conceptual requirements for presentation at the International Team workshop in October 2000. It was understood, however, that this was only a temporary solution and that the case study process required some adjustment to achieve better results from the diplomatic analysis in subsequent rounds. Accordingly, two changes were made to the process of conducting case studies.

The first change concerned the kinds of systems that would be targeted in subsequent rounds. Given that the majority of systems that the Authenticity Task Force had examined thus far had not contained records when viewed from the perspective of contemporary diplomatics, Task Force Researchers were faced with two choices--they could either revise the eligibility criteria for treating the entities within electronic systems or the electronic systems themselves as records to accommodate the various dynamic realities they were seeing; or they could circumscribe the range of case studies to accommodate only those systems that contained entities that fit the diplomatic construct of a record. The researchers opted for the latter route on the grounds that one of the reasons for choosing diplomatics as a means of analyzing electronic records was to evaluate its effectiveness. The researchers needed to examine a range of systems that fit the construct in general terms before they could evaluate its effectiveness in more specific terms. Accordingly, the case study selection criteria were adjusted to ensure that only electronic systems containing, or having the potential to contain, records were selected in subsequent rounds.[30]

The second change concerned the designation of responsibility for preparing the *TEDGI*s. It was decided that the UBC research assistants would prepare the *TEDGI*s because their knowledge of diplomatics made them the best equipped to map the answers to questions on the *CSIP* to the relevant archival-diplomatic element of the *TEDGI*. Once the research assistants had completed the *TEDGI*s, they were required to send them back to the case study interviewers for verification of the accuracy of the mapping before they were finalized. The *TEDGI*s subsequently became the basis for the preparation of draft versions of the diplomatic analyses of case studies, which were also returned to the case study interviewers for approval prior to being finalized. It was decided also that the experiment of developing hypothetical scenarios would not be repeated in the next rounds. While the exercise had helped the research assistants to

[29] The procedural rules may be found on the website of the UBC project at <http://www.interpares.org/UBCProject/index.htm>.
[30] See above, fn. 21.

understand how authenticity requirements might manifest themselves in a given situation, the International Team found them to be confusing and overly prescriptive.

Apart from these changes, the process of diplomatic analysis in the second and third phases (rounds three and four) was similar to the process in the first phase (rounds one and two). The main difference was that case studies from rounds three and four contained more systems with records.

Twenty-two case studies were analyzed from an archival diplomatic perspective. Of these, twelve systems were found to contain records. In the systems containing records, many of the elements associated with uniquely identifying a record and placing it in its immediate context were either implicit or absent. For example, in most of the systems there was no explicit manifestation of the archival bond between and among the records participating in the same action. Moreover, while it was reasonably straightforward to identify the business processes supported by the electronic system in general terms, it was not always easy to determine how the records participated in or supported specific actions.

Authenticity Task Force researchers had hypothesized at the outset of the research that intrinsic and extrinsic elements of documentary form and annotations would play key roles in establishing the identity and demonstrating the integrity of electronic records. This hypothesis failed to be supported, however, by either the diplomatic analysis or the analysis of elements relating to the identity and integrity of records described in 3.2.3.3(i). In the case studies analyzed, it was often difficult to determine the significance of the presence or absence of annotations or specific elements of documentary form. The determination of documentary forms in general and the establishment of required elements of form in particular appeared to be deeply embedded within specific institutional and procedural contexts and were resistant to any easy generalizations. As a result, the researchers were unable to draw any general conclusions about the relevance of specific intrinsic and extrinsic elements of documentary form or annotations to a consideration of an electronic record's authenticity outside of the specific institutional and procedural context in which the record was created (this is discussed further in section 3.2.3.2).

At the same time, however, it was possible to identify certain commonalities in the means used by creators to protect record authenticity from one institution to the next. The diplomatic analysis and the analysis of elements relating to identity and integrity revealed that record creators tend to rely on procedural means for protecting authenticity and to treat it as part of the management of the electronic system as a whole rather than as part of the management of individual records within the system. The commonest means identified were access privileges (including passwords, user IDs, user profiles), followed by the use of audit trails and backup procedures.

3.2.3.2. Development of a Typology of Electronic Records

The diplomatic analyses of case studies were undertaken to facilitate the identification of general conceptual requirements for authenticity and the development of a typology of electronic records based on authenticity requirements for specific types of electronic records. The primary purpose of any typology is "to produce ordered and reproducible sets that can support the rapid identification of members of groups of sets in general and members of individual sets or subsets in particular."[31] The design and implementation of a typology may be approached from the top down or the bottom up. As Seamus Ross explains:

> In the former approach a researcher begins within the premise that a 'group of entities' …forms a bounded set. Then the researcher attempts to select and define characteristics shared by the material and to determine whether objects/entities proposed as members of the group have the required attributes. In this approach the set becomes equivalent with the type. In the second approach the investigator starts with the objects and proceeds to describe the component elements. The elements are then grouped into attributes and the attributes subsequently grouped into restricted sets. These are shared component types that carry meaning.[32]

The criterion for developing the typology of electronic records was the significance of specific extrinsic and intrinsic elements of documentary form and annotations for carrying out or attesting to the action or matter in which a record participated. Between October 2000 and April 2001, the Task Force explored numerous candidate types based on a range of criteria. A top-down approach was adopted for the identification of these types, mainly because there were insufficient data from the case studies to support a bottom-up approach.

The initial basic typology reflected the four categories of records identified by contemporary archival diplomatics, based on the relationship between a record and the action in which it participates. This categorization was chosen on the grounds that groups of records sharing the same function with respect to an action or matter form a bounded set. The categories are *dispositive* records (records whose written form is required by the juridical system as the essence and substance of an action), *probative* records (records whose written form is required by the juridical system as proof that an action has taken place prior to its documentation), *supporting* records (records whose written form is discretionary; they are created to provide support for, and are procedurally linked to, an action), and *narrative* records (records whose written form is also discretionary; they do not participate procedurally in the action but are created as part of the process of

[31] Seamus Ross, "Dress-pins from Anglo-Saxon England: their production and typo-chronological development," D.Phil. dissertation, University of Oxford, 1992, 68.
[32] *Ibid*., 86.

setting oneself to work). These categories were extended by the Task Force to refer to the smallest indivisible aggregation of records (for example, the file unit) in each system rather than to individual records. This definitional extension of the record categories implied an extension of the authenticity requirements because the requirements for a given category of record aggregation (dispositive, probative, supporting, narrative) would apply to all the records within the aggregation, regardless of the different types of individual records contained within it.

The Task Force hypothesized that for dispositive and probative aggregations of records, that is, records whose written form is required, the elements of extrinsic and intrinsic form as well as annotations would be prescribed by the juridical system and, therefore would have to be preserved in their entirety; whereas for supporting and narrative aggregations of records, that is, records whose written form is not required, it was assumed that there would not be the same necessity to preserve all the elements and annotations. This hypothesis was not, however, supported by the case studies which suggested that: (1) the requirement of a written form does not necessarily translate into specific required elements of documentary form or annotations; and (2) the fact that a written form is not required, does not necessarily translate into an absence or reduction of specific required elements of documentary form or annotations since there are cases of supporting and narrative records whose written form is highly regulated.

Next, the Task Force explored the possibility of a typology of electronic records based on the diplomatic categorization of procedures. These include: *constitutive* procedures (procedures which create, extinguish, or modify the exercise of power and which may be further subdivided into procedures of concession, of limitation, or of authorization); *executive* procedures (procedures which allow for the regular transaction of affairs according to rules established by an external authority); *instrumental* procedures (procedures connected to the expression of opinions and advice); and *organizational* procedures (procedures whose purpose is to establish organizational structure and internal procedures and to maintain, modify, or extinguish them). Since the categories of procedure imply different levels of documentary control, with constitutive procedures being the most controlled and instrumental procedures being the least controlled, the Task Force hypothesized that records created in accordance with the more controlled procedures would have more required elements of documentary form and annotations than would those created in accordance with less controlled procedures. This categorization was ultimately rejected, however, on the grounds that (1) records do not necessarily aggregate in accordance with these procedures; and (2) it is not possible to generalize, simply on the basis of the procedure, about the significance of elements of documentary form and annotations.

The Task Force experimented with a number of other candidate types, based on a range of criteria, such as whether the system contained records or was itself a record; whether the system contained one type of records, or many types; whether the records were digital or digitized, and so on. None of these types, however, resulted in a categorization of records on the basis of which specific requirements for authenticity could be formulated.

In April 2001, the Task Force had not yet succeeded in developing a typology that provided a meaningful differentiation and specification of requirements for authenticity according to types of records.[33] Despite our efforts, we were simply unable to establish a correlation between authenticity and the presence of specific documentary elements or annotations. Since the deadline for submitting the final version of the requirements for authenticity was June 2001, the Task Force decided to suspend its efforts to develop a typology and to focus instead on refining the general conceptual requirements for assessing the authenticity of electronic records.

3.2.3.3. Additional analyses of case studies

In addition to the diplomatic analysis, and in order to support the Authenticity Task Force's theory-building efforts, four other types of analysis were performed at UCLA and the University of Albany on some or all of the case studies.

Prior to these analyses, *TEDGI* and selected *CSIP* data were entered into a database and interview tapes, where available, were transcribed.

A preliminary analysis of completed TEDGIs and supporting documentation was undertaken in order to:

1) create a narrative overview for each case study
2) generate tables of how each TEDGI element was completed across case studies
3) verify how each TEDGI element was completed and attempt to reconcile any differences between the interviewer completing the TEDGI in the first two rounds and the research assistants analyzing the data
4) identify which questions were used to support completion of which elements, and which questions were seldom, if ever used

[33] Researchers working on the D(igital) A(rchivering in V(laamse) I(nstellingen en) D(iensten) Project in Brussels reached a similar conclusion. The original aim of the DAVID Project "was to work out a typology from which a method for preserving the various types of digital archive documents over the long term would follow." According to the researchers, "[t]his typology would stand or fall on its usefulness in formulating a preservation strategy, and was pursued with this goal in mind. The first attempt rested on the editorial form and function of the digitally preserved document, a method of description and classification borrowed from paper archiving. It was soon obvious, however, that this was no basis for managing digital archives and no basis for formulating a preservation strategy." See Filip Boudrez, "The Digital Recordkeeping System: Inventory, Information Layers, and Decision-Making Model as Point of Departure," (Antwerp, June 2001), 4, at <http://www.antwerpen.be/david>.

5) identify what supporting documentation was used to complete which elements
6) identify elements that interviewers had difficulty completing, or where there appeared to be little consistency in how they were completed
7) track the numbers of interviews and the position titles of interviewees necessary to complete case studies

A detailed analysis of completed TEDGIs and supporting data was then undertaken to identify:

1) What are the elements that are most commonly present across case studies?
2) What are the elements that are most commonly absent, or cannot be discerned across case studies?
3) What are the business functions being supported by the electronic systems studied?
4) What are the activities and transactions performed by the electronic systems in support of the business functions?
5) At which level within the organization do the electronic systems exist?
6) What are the relationships between paper and electronic components of record-keeping systems?

A narrative analysis of selected transcribed interviews was also undertaken to identify:

1) In what ways do records creators, custodians, and systems personnel conceptualize the nature and role of the electronic records and/or record-keeping system being studied?
2) What are the variances in language used to describe records by records creators, custodians, and systems personnel?
3) The extent to which the findings of 1) and 2) should or could be factored into the design of a method to identify and ensure the preservation of authentic electronic records.

As outlined above, upon commencing the data analysis, researchers first created a brief narrative description of the case study, referred to as the case study overview, based upon the documentation submitted for analysis by the interviewers. The draft overview was then returned to the interviewers for them to review together with the respondents and make any necessary corrections that might be due to misinterpretation of the case study data.

The researchers then proceeded to undertake the following four analytical activities:

i) *Analysis of how and to what degree the identity and integrity of electronic records is supported within and across case studies*. In undertaking the diplomatic analysis of the case studies, the researchers had begun with an assumption that the diplomatic elements of electronic records would be the same (or at least the fundamental elements would be similar) as those of traditional records. However, researchers began to realize that these elements are less

explicit in electronic records, and that more of the record's identifying elements are found in its context, instead of on the face of the records, as was the case for traditional records. As a result, the diplomatic analysis often focussed on what was wrong with the systems that were studied when held up against the ideal record represented by the *Template*, rather than effectively identifying alternative, new, or unanticipated ways in which authenticity requirements were being met in these systems. In order to facilitate the Task Force's efforts to develop a typology of authenticity requirements for electronic records, therefore, each case study was analyzed in order to determine which, if any aspects of the systems examined corresponded to, or supported elements establishing the identity and integrity of electronic records (the key concerns of authenticity) as delineated in the *Template for Analysis*. This analysis examined not only specific elements, but a variety of contexts, sources and techniques through which elements might be manifested or their purposes achieved.

The case study data was coded to see whether any patterns were discernible, across all case studies, or across those that seem likely to contain similar types of records. The resulting analysis showed that within individual, and across case studies authenticity is assured mainly through procedural means and treated as part of the management of the electronic system as a whole.

ii) *Characteristics of case studies by type of information system.* This analysis applied a model commonly used in business administration to identify types of information systems developed and used in an organization to support business processes and to fulfil the mission of the organization.[34] The model provided one way to describe the nature of systems found in an organization, and, thereby, potentially a method to help discern systems that are likely to create records, and whether those records are likely to be dispositive, probative, supporting, or narrative.

In this model, an organization is divided into four levels:

1) *Operational-level systems*: information systems that monitor the elementary activities and transactions of the organization.

2) *Knowledge-level systems:* information systems that support knowledge and data workers in an organization

3) *Management-level systems:* information systems that support the monitoring, controlling, decision-making, and administrative activities of middle managers

4) *Strategic-level systems:* information systems that support the long-range planning activities of senior management.

[34] Kenneth C. Laudon and Jane P. Laudon, *Management Information Systems: New Approaches to Organization and Technology* (New Jersey: Prentice Hall, 1996)

Organizational functions are supported by six major types of systems:

1) *Transaction processing system (TPS):* computerized system that performs and records the daily routine transactions necessary to conduct the business; these systems serve the operational level of the organization

2) *Knowledge work system (KWS):* information system that aids knowledge workers in the creation and integration of new knowledge in the organization.

3) *Office automation system (OAS):* Computer system, such as word processing, electronic mail system, and scheduling system, that is designed to increase the productivity of data workers in the office

4) *Management information system (MIS):* information system at the management level of an organization that serves the functions of planning, controlling, and decision making by providing routine summary and exception reports.

5) *Decision-support system (DSS):* information system at the management level of an organization that combines data and sophisticated analytical models to support semi-structured-decision making.

6) *Executive Support System (ESS):* information system at the strategic level of an organization designed to address unstructured decision making through advanced graphics and communications.

Operational level systems such as transaction processing systems help operational managers keep track of the organization's everyday activities. Knowledge level systems such as office automation systems and knowledge work systems help knowledge and data workers design products, distribute information and manage paperwork. Management level systems such as management information systems and decision support systems help middle managers monitor and control business activities. Strategic level systems such as executive support systems help senior managers with long-term planning. The model also delineates the information inputs, processes, and outputs that serve as indicators of the type of system being examined.

This approach closely parallels certain traditional appraisal approaches that have targeted executive and administrative levels within an organizational hierarchy as being most likely to generate key records relating to policy, procedural and organizational decision-making[35]. In the model used in this analysis, the types of information systems commonly associated with these levels would be management information systems (MIS), Decision Support Systems (DSS), and Executive Support Systems (ESS).

[35] See, for example, Schellenberg, T.R. *Modern Archives: Principles and Techniques* (Chicago: Society of American Archivists, 1998): 142-143

This analysis examined the organizational level and information inputs, processes, and outputs associated with each case study in order to try to identify the type and nature of each system and the likelihood that it generates, or should generate, records. In the analysis, in recognition of the "mixed" nature of most of the systems studied, the researchers also extended the model to identify more closely both electronic and paper outputs. Because stable content is considered to be an identifying characteristic of authentic records, researchers further categorized the status of system outputs in order to understand the degree to which they were stable:

- Fixed: Once output is created, it is immutable. If it needs to be changed, either an update must be appended, or a new version must be created.

- Transient: Output is created for temporary use only, for example, a screen display providing the results of an information query.

- Dynamic: Output is stored on the system but can be changed, updated, annotated, and overwritten.

The analysis indicated the complexity of the systems studied--almost no system exists independent of a wider record-keeping system, and most relate to more than one organizational level and perform a range of functions rather than conforming to one of the discrete types contained in the business model. Equally, most of the systems studied have components that are paper as well as those that are electronic. This "mixed" environment must be taken into account when understanding the nature of any potential record generated by the system. Many of the systems studied contained primarily transactional data, and most of them generated primarily transient or dynamic output.

The majority of the case studies focussed on systems that function at the operational and knowledge levels within the organization, and less frequently at the management level. In comparing this analysis with the diplomatic analysis of the same case studies, one can see that those systems identified through the diplomatic analysis as containing, or that should contain dispositive or probative records for the most part carry out at least some management as well as operational and knowledge-level functions. One could speculate, therefore, that systems addressing functions at the management level and above would be more likely to contain records and less transactional data.

iii) *Functional analysis of case studies.* As the research progressed, it became increasingly clear that understanding the nature and boundaries of electronic records required a detailed understanding of the business functions and activities of the record-keeping systems being studied. Researchers at UCLA selected the method delineated in the National Archives of Australia's *DIRKS (Designing and*

Implementing Recordkeeping Systems) Manual[36] as one of the most robust and replicable extant approaches to functional analysis. The purpose of conducting this functional analysis was to describe, unambiguously for non-archivists, and systems designers in particular, the nature of the record-keeping function performed by the system. The researchers concluded, however, after attempting both a narrative and graphical representation of the major functions of the systems being studied, and a breakdown of the actions and transactions that support those functions, and then receiving feedback from interviewers and respondents upon the draft breakdowns, that it was not possible to render an accurate functional decomposition of each case study. The reason for this was because the *CSIP*, developed from the diplomatic perspective of analyzing individual documents, had not been designed to capture the appropriate depth of functional detail about the record-keeping system as a whole.

iv) *Narrative analysis of transcribed case study interview data.* One concern of the researchers was that their understanding of the nature of electronic records and the concept of authenticity, and how that understanding was expressed through the terminology used in the *Case Study Interview Protocol* and any InterPARES products, would not match that of, or be understandable by, record-keepers and systems personnel. Although the case study interviews were heavily scripted to ensure some level of consistency across cases, some interviews were recorded and transcribed (where respondents gave their permission) and of these interviews, some contained additional discussion about the nature and functionality of the electronic record-keeping in which the respondents were engaged. Selected case study transcripts were examined to gain a closer understanding of respondent perspectives and terminology. A complete narrative analysis was conducted of one case study that demonstrated, even though the process of transcribing and analyzing interview data is laborious and time-consuming, the value of such an approach for future research.[37] It should be noted, however, that the case studies were not originally intended to be subjected to narrative analysis. Had this been the case, interviews, or components of interviews would need to have been conducted in a more free-form or conversational manner which would allow respondents to expand their commentary and which would avoid providing respondents with InterPARES' own terminology and rhetorical tropes.

4. RESEARCH FINDINGS

[36] National Archives of Australia, *Designing and Implementing Recordkeeping Systems: Manual for Commonwealth Agencies*. Available at: <http://www.naa.gov.au/recordkeeping/dirks/dirksman/dirks.html>.

[37] See Ciaran Trace, "Applying Content Analysis to Case Study Data: A Preliminary Report," available on the InterPARES website.

4.1 Preamble

The purpose for developing the *Template for Analysis* and testing its effectiveness through four rounds of case studies was to lay the foundation for establishing conceptual requirements for assessing and maintaining the authenticity of electronic records over the long-term. The requirements are described in detail in *Requirements for Assessing and Maintaining the Authenticity of Electronic Records* (see Appendix to this report) and embody the major conceptual findings of the Authenticity Task Force.

4.2 Conceptual findings: the requirements for authenticity

4.2.1 Terms of assessment of authenticity

To assess the authenticity of an electronic record, the preserver must be able to establish its *identity* and demonstrate its *integrity*.

The *identity* of a record refers to the distinguishing character of a record, that is, the attributes of a record that uniquely characterize it and distinguish it from other records. From an archival-diplomatic perspective, such attributes include: the names of the persons concurring in its formation (that is, its author, addressee, writer, and originator); its date(s) of creation (that is, the date it was made, received, and set aside) and its date(s) of transmission; an indication of the action or matter in which it participates; the expression of its archival bond, which links it to other records participating in the same action (for example, a classification code or other unique identifier); as well as an indication of any attachment(s) since an attachment is considered an integral part of a record.

The *integrity* of a record refers to its wholeness and soundness: a record has integrity when it is complete and uncorrupted in all its essential respects. This does not mean that the record must be precisely the same as it was when first created for its integrity to exist and be demonstrated. Even in the paper world, with the passage of time, records are subject to deterioration, alteration and/or loss. In the electronic world, the fragility of the media, the obsolescence of technology and the idiosyncrasies of systems likewise affect the integrity of records. When we refer to an electronic record, we consider it essentially complete and uncorrupted if the message that it is meant to communicate in order to achieve its purpose is unaltered. This implies that its physical integrity, such as the proper number of bit strings, may be compromised, provided that the articulation of the content and any required elements of form remain the same.

4.2.2 Assessment and Maintenance of Authenticity

The preserver must assess the authenticity of electronic records before electronic records are transferred to archival custody and maintain it after transfer. The assessment is an integral part of the records' appraisal while the maintenance is an integral part of their long-term preservation.

Before records are transferred to archival custody it is necessary for the preserver to establish, as part of the process of appraisal, whether and to what extent the records have been maintained by the creator using technologies and administrative procedures that either guarantee their authenticity or at least minimize risks of change from the time the records were first set aside to the point at which they are subsequently accessed.

After the authenticity of the creator's electronic records has been established in the appraisal process, and the records transferred from the creator to the preserver, their authenticity needs to be maintained by the preserver. To do so, the preserver must maintain the electronic records in accordance with procedures that ensure their continuing authenticity and produce copies of those records in accordance with procedures that ensure that their authenticity is not compromised by the reproduction process. To support its attestation of the authenticity of copies of electronic records, the preserver must also produce and maintain documentation relating to the manner in which it has maintained the records over time as well as the manner in which it has reproduced them.

In light of the above, the Authenticity Task Force has developed two sets of requirements: the first set includes requirements that support the presumption of the authenticity of the creator's electronic records before they are transferred to the custody of the preserver, while the second group includes requirements that support the production of authentic copies of electronic records that have been transferred to the custody of the preserver. The first set of requirements are termed "benchmark requirements" while the second set are termed "baseline requirements."

4.2.3 Conceptual framework of the benchmark and baseline requirements

Both the benchmark and the baseline requirements are based on the notion of trust in record-keeping and record preservation. The benchmark requirements draw specifically on the notion of a *trusted record-keeping system*, while the baseline requirements are predicated on the role of the preserver as a *trusted custodian*.

A *trusted record-keeping system* has been defined as "a type of system where rules govern which documents are eligible for inclusion in the record-keeping system, who may place records in the system and retrieve records from it, what

may be done to and with a record, how long records remain in the system, and how records are removed from it."[38]

The *role of the preserver as trusted custodian* dates back to Roman antiquity when citizens would deposit private records in the Tabularium for the express purpose of rendering them authentic. As a trusted custodian of records, ancient archival institutions sustained and lent credibility to contractual relationships between citizens. They also lent credibility to the implicit social contract between citizens and the state by preserving the records of the state's past actions on the basis of which the state could be held to account. Today, the role of the preserver as a trusted custodian is also analogous to that of the trusted third party record-keeper in electronic contracting. A trusted third party record-keeper is a physical or juridical person entrusted with independently maintaining the records of Electronic Data Interchange (EDI) partners. The reason for having a trusted third party record-keeper is to increase the probability that records of an EDI transaction will be accepted in court as evidence. To be considered a trusted record-keeper, the person must demonstrate, among other things, that it has no reason to alter retained records itself; that it has no interest in allowing others to alter records; and that it is capable of implementing security procedures to a degree that meets the necessary standards of integrity and accuracy.[39] Similarly, to be considered a trusted custodian, the preserver must demonstrate that it has no reason to alter the preserved records, or to allow others to alter them, and that it is capable of implementing the baseline requirements.

4.2.4 Specific conceptual framework for the benchmark requirements for assessing the authenticity of the creator's electronic records

The creator's records belong to one of two categories. The first category comprises those records that exist as created. They are considered authentic because they are the same as they were in their first instantiation. The second category comprises those records that have undergone some change and therefore cannot be said to exist as first created; they are considered authentic because the creator treats them as such by relying on them for action or reference in the regular conduct of business. However, the authenticity of electronic records is threatened whenever they are transmitted across space (that is, when sent to an addressee or between systems or applications) or time (that is, either when they are in storage, or when the hardware or software used to store, process, or communicate them is updated or replaced). Given that the acts of setting aside an electronic record for future action or reference and of retrieving it inevitably entail moving it across significant technological boundaries (from display to storage subsystems and vice versa), virtually all electronic

[38] Margaret Hedstrom, "Building Record-Keeping Systems: Archivists Are Not Alone on the Wild Frontier," *Archivaria* 44 (Fall 1997): 57

[39] Bernard D. Reams Jr., L.J. Kutten, and Allen E. Strehler. *Electronic Contracting Law: EDI and Business Transactions, 1996-97 Edition* (New York: Clark, Boardman, Callaghan, 1997), 37.

records belong to the second category. Therefore, the preserver's inference of the authenticity of electronic records must be further supported by evidence – provided in association with the records – that they have been maintained using technologies and administrative procedures that either guarantee their continuing identity and integrity or at least minimize risks of change from the time the records were first set aside to the point at which they are subsequently accessed. The requirements for assessing the authenticity of the creator's electronic records concern this evidence.

4.2.4.1 The presumption of authenticity

A presumption of authenticity is an inference that is drawn from known facts about the manner in which a record has been created, handled, and maintained. The evidence that supports the presumption that the creator created and maintained its electronic records authentic are enumerated in the *Benchmark Requirements Supporting the Presumption of Authenticity of Electronic Records* (Requirement Set A). A presumption of authenticity will be based upon the number of requirements that have been met and the degree to which each has been met. The requirements are, therefore, cumulative: the higher the number of satisfied requirements, and the greater the degree to which an individual requirement has been satisfied, the stronger the presumption of authenticity. This is why these requirements are termed 'benchmark' requirements.

4.2.4.2 The verification of authenticity

In any given case, there may be an insufficient basis for a presumption of authenticity, or the presumption may be extremely weak. In such cases, further analysis may be necessary to verify the authenticity of the records. A verification of authenticity is the act or process of establishing a correspondence between known facts about the record and the various contexts in which it has been created and maintained, and the proposed fact of the record's authenticity. In the verification process, the known facts about the record and its contexts provide the grounds for supporting or refuting the contention that the record is authentic. Unlike the presumption of authenticity, which is established on the basis of the benchmark requirements, this verification involves a detailed examination of the records themselves and reliable information available from other sources about the records and the various contexts in which they have been created and maintained. Methods of verification include, but are not limited to, a comparison of the records in question with copies that have been preserved elsewhere or with backup tapes; comparison of the records in question with entries in a register of incoming and outgoing records; textual analysis of the record's content; forensic analysis of aspects such as medium and script; a study of audit trails; and the testimony of a trusted third party.

4.2.5 Specific conceptual framework for the baseline requirements supporting the production of authentic copies of electronic records

After the records have been presumed or verified authentic in the appraisal process, and have been transferred from the creator to the preserver, their authenticity needs to be maintained by the preserver. In order to do so, the preserver must carry forward the records in accordance with the baseline requirements that apply to the maintenance of records, producing copies according to procedures that also maintain authenticity. The production of authentic copies is regulated by the *Baseline Requirements for the Production of Authentic Copies of Electronic Records* (Requirement Set B). Unlike the Benchmark Requirements, all of the requirements included in the Baseline Requirements must be met before the preserver can attest to the authenticity of the electronic copies in its custody. This is why the requirements for the production of authentic electronic copies are termed 'baseline' requirements.

Satisfaction of these baseline requirements will enable the preserver to certify that copies of electronic records are authentic. Traditionally, the official preserver of the records has been the person entrusted with issuing authentic copies of such records. To fulfill that role, the preserver needed simply to attest that the copy conformed to the record being reproduced. With electronic records, the difficulties related to preservation make it prudent for the preserver to produce and maintain documentation relating to the manner in which it has maintained the records over time as well as the manner in which it has reproduced them to support its attestation of authenticity.

A copy is the result of a reproduction process. A copy can be made from an original or from a copy of either an original or another copy.[40] There are several types of copy. The most reliable copy is a 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.

[40] In common language, *copy* and *reproduction* are synonyms. For the purposes of this research, the term reproduction is used to refer to the process of generating a copy, while the term copy is used to refer to the result of such a process, that is, to any entity which resembles and is generated from the records of the creator. An original record is defined as the first, complete record, which is capable of achieving its purposes (that is, it is effective). A record may also take the form of a draft, which is defined as a temporary compilation made for purposes of correction. For a discussion of the status of originals, drafts, and copies in an electronic environment see Luciana Duranti and Heather MacNeil, "The Protection of the Integrity of Electronic Records: An Overview of the UBC-MAS Research Project." *Archivaria* 42 (Fall 1996): 56-57. For the definition and interpretation of an original in the context of international law and electronic commerce, see United Nations Commission on International Trade Law, *UNCITRAL Model Law on Electronic Commerce with Guide to Enactment* (New York: United Nations, 1997), esp. article 8 of the "Model Law" and para. 62-69 of the "Guide to Enactment".

Any of these types of copy is authentic if attested to be so by the official preserver. By virtue of this attestation, the copy is deemed to conform to the record it reproduces until proof to the contrary is shown. Such attestation is supported by the preserver's ability to demonstrate that it has satisfied the applicable baseline requirements for maintenance and all of the requirements for the production of authentic copies.

4.3 Methodological findings

While the primary purpose of the work of the Authenticity Task Force has been to address authenticity requirements for electronic records, a significant by-product of its work, and indeed that of the entire InterPARES Project, has been an enhancement and extension of existing archival methodological knowledge and expertise. Drawing upon the multi-disciplinary expertise of its researchers, InterPARES applied a diverse range of theoretical and applied approaches, including diplomatic analysis, modeling, and narrative analysis. This diversity of approaches was unprecedented in archival research to date, and throughout its work, the Authenticity Task Force strived to assess and document what worked in the different methods that it used, what partially worked, and what did not, and why.

4.3.1 Limitations of diplomatics as an analytical tool

One reason for incorporating the perspective of contemporary archival diplomatics into the work of the Authenticity Task Force was to evaluate its effectiveness as an analytical tool. The Authenticity Task Force found it to be a useful means of assessing the strengths and weaknesses of current electronic systems. For example, it highlighted the extent to which electronic systems are still being designed to manage data rather than records. This appears to be the case even when the purpose for which the system is designed would appear to require the creation and maintenance of fixed records rather than fluid data. It also highlighted the significant extent to which elements relating to a record's identity are implicit rather than explicit; and the consequent need to make certain identifying elements explicit to ensure that knowledge of key indicators of identity is not lost when the records are removed from the system in which they have been created and actively used. Finally, the diplomatic analysis revealed a surprising level of indifference on the part of record creators to authenticity-related issues, an indifference attributable mainly to a (possibly misplaced) confidence in the capacity of generic technological and procedural controls over the electronic system to protect the authenticity of the records contained within it.

At the same time, contemporary archival diplomatics, as currently articulated, remains rooted in a very traditional conception of what a record is and is thus limited in its capacity to extend the range of archival understanding about the nature of different kinds of electronic systems and the variety of entities contained within them. While it is quite effective in analyzing electronic

environments that are analogous to traditional record-keeping environments, it is considerably less helpful in analyzing electronic environments that are not so analogous. This finding points to the limits of the known as an aid to understanding the unknown. To increase the utility of diplomatics as an aid to understanding diverse electronic systems, it will be necessary to develop a more nuanced interpretation of the characteristics of electronic records and the manner in which they manifest themselves in a variety of electronic environments. While the Authenticity Task Force began to move in this direction in the final two rounds of case studies, where it focused less attention on establishing whether the record was complete, stable, and unchangeable, and more attention on determining whether and to what extent the system was capable of tracking and preserving any changes, considerably more interpretive work is needed.

The limitations of the diplomatic model of a record as it is elaborated in the *Template for Analysis* are attributable mainly to the fact that the model was built on the premises of *general diplomatics*. *General diplomatics* seeks to decontextualize records, to eliminate their particularities, variations and anomalies in the interest of identifying the common, shared elements of records that cut across juridical, provenancial, and technological boundaries. Given the complexity and variety of electronic systems it might make more sense to adopt and adapt the approach of *special diplomatics*, which, traditionally, has focused on the records of individual chanceries and specific juridical systems (the typology of papal chancery documents prepared in the course of the Authenticity Task Force's work is an example of special diplomatics). In such an approach, one would begin with an analysis of the various features of the systems themselves and the broader record-keeping environment in their own terms, with all their particularities, variations, and anomalies; and, on the basis of that analysis begin to build a more general framework.

Further refinement of the diplomatic approach is also needed to accommodate record aggregates. One significant difference between the diplomatic and the archival perspective is that diplomatics focuses mainly upon the individual document or record, while archival science tends to emphasize the record aggregate (for example, the fonds or the series). Although researchers attempted, during the development of the *Template* to incorporate the aggregate approach, the *Template* remained predominantly focused on elements that are only relevant at the level of individual records. Many of the systems examined through the case studies, however, contained heterogeneous aggregates of records. In fact, the archival extensions of the *Template*, such as the addition of the five categories of context (juridical-administrative, provenancial, procedural, technological, documentary), turned out to be the most relevant to an understanding of the record-keeping environment, and the grounds on which creators based their presumption of the records' authenticity. These contexts were, however, the least well developed part of the *Template*. For example, in several case studies, audit trails were identified by the creator as a significant means of ensuring the authenticity of electronic records. Audit trails are part of

system administration and therefore were considered an element within the record's technological context. The element "system administration" was not decomposed sufficiently, however, to enable Task Force researchers to identify the various kinds of audit trails and the specific purposes they serve in a given environment. In the absence of that identification, it was difficult to assess the extent to which an audit trail supported the creator's presumption of authenticity in particular cases.

To deal more effectively with such systems, therefore, the researchers believe that a contemporary archival diplomatic analysis should seek to identify and elaborate more completely the nature of archival aggregates and the elements that uniquely characterize them.

4.3.2 Limitations of case study design and instrumentation

One objective of the case studies was to make recommendations about the development of procedures, instrumentation, and analytical techniques to assist archivists and records managers in gathering and assessing the information they need in order to identify and preserve authentic records in electronic systems. While the Task Force found that the case study method was a very valuable approach to understanding the nature of the electronic record, the analysis of the successive rounds of case studies pointed up several areas where the design and instrumentation of the case studies were problematic or could effectively be refined. Based on a tandem evaluation of instrumentation and method, after the analysis of each round of the case studies, the instrumentation and protocol were modified accordingly and tested through implementation in the succeeding round.

Some of the issues that arose with the case studies included the following:

- The contact phase, which consisted of getting permission from the relevant institution and administrators to conduct the case study, and then identifying the appropriate respondents, prior to conducting the actual interview or interviews was often lengthy. This issue, combined with the time it took to revise and get human subjects' approval for the case study protocol, difficulties scheduling the interviews, the interviewers' time to complete the TEDGI, and the time it took to analyze the resulting data all contributed to difficulties in keeping to the tight schedules the Task Force had identified for each round of case studies. It is likely that a case study approach would also be time-consuming for practicing archivists to implement when studying their institution's records.

- The *Case Study Interview Protocol(CSIP)* was too long. It is a daunting instrument, both for interviewers and respondents. Notwithstanding this, a three-hour interview is generally insufficient to cover all the questions in detail and follow-up contacts are time and labor intensive for all parties concerned. Gathering sufficient information on electronic record-keeping, whether for a

research project or for institutional records management and archival purposes is necessarily time-consuming and complex. Researchers, practitioners, and record-keepers should not under-estimate the resource-intensiveness of information gathering.

- Many questions from the *CSIP* were never used in populating the *TEDGI* and could potentially be eliminated from the *CSIP*, thus streamlining it further.

- The sometimes arcane terminology in the *CSIP*, drawing upon that of diplomatics, was not sufficiently oriented towards that used or understood by the people who are being interviewed. Moreover, some of the same terminology is used, but with different meanings, by the systems community. While the establishment of a project glossary has sought to address these issues, they still potentially impeded the interview process with implications for the reliability of some of the answers obtained in the case studies.

- Making the translation between the *CSIP* and the *TEDG*I is difficult for anyone not trained in diplomatics even with the explanations provided with the *Template for Analysis*. Moreover, the correlation between the *CSIP* and *Template Element Data Gathering Instrument (TEDGI)* was not always clear. In many instances, the interviewers making the translation between the two instruments often got confused or simply made errors, thus potentially affecting the reliability of the data. Inconsistencies in how the *TEDGI* was completed became evident when completed *TEDGI*s were compared across case studies, requiring that the researchers analyzing the data go back to the *CSIP*, and sometimes also the interviewers, to verify how the TEDGI was completed. The translation process was also extremely time-consuming. In the final two rounds of the case studies these issues were addressed by having the *TEDGI* completed from the *CSIP* by researchers at the University of British Columbia, rather than by individual interviewers.

- In situations where case study interviewers were also practicing archivists, they brought to bear valuable experience and institutional knowledge, as well as archival expertise. Without such knowledge, some nuances of the records environment might be missed. The limitation of such situations, however, is that the interviewers have a considerable amount of implicit and unconscious knowledge, because of their familiarity with the institution and its records, that is not always captured overtly in case study data (especially the translation of the data into the *TEDGI*). The same would potentially be true if the interviewer came from a systems background. An alternate, although more labor-intensive approach might be to use pairs of interviewers with different backgrounds to conduct the interviews.

- It was often unclear what the focus of or unit of analysis for the case study was supposed to be – was it the entire record-keeping environment (that is, the mandate, the business processes, the data input and output, whether

electronic or paper) or the electronic system alone? Not all electronic records systems can be analyzed and managed at the same level of granularity. Sometimes it is possible to examine records document by document (for example, with e-mail). In other cases, one has to approach an entire system (for example, with databases). In yet other cases, a single record aggregate comprises both paper and electronic components. The most viable approach appears to be to commence with a thorough understanding of the record-keeping environment and allow that understanding to delimit the types of records that might be present and their intellectual and physical parameters. The *CSIP*, however, was weak in terms of collecting data that will allow for the analyses or understanding of specific record-keeping functions and events. This made a functional analysis of the cases studied, as well as the development of a typology based upon specific record-keeping acts or functions difficult to achieve.

- Because it was first derived from diplomatics and what is known about traditional records, the *CSIP* unconsciously favored record-keeping systems that look most like their paper counterparts. This made it difficult to understand whether researchers were not finding specific diplomatic elements because they were absent or because they were not aware of how those elements might be manifested differently in electronic systems. It also made it difficult to identify whether elements were absent because the form or the record has changed or because the case studies were examining imperfect record-keeping systems that were not generating or maintaining good records.

Based upon the Task Force's analysis of the quality and scope of the case study data, it makes the following recommendations for revisions to the design and instrumentation of record-keeping case studies:

1) Explicitly examine the entire record-keeping system, and not just its electronic components.

2) Re-orient the *CSIP* to start with an analysis of business processes - proceeding from the general to the specific and delineating functions, activities, and then transactions. This will make it easier to identify actions in which records participate, and the nature of that participation.

3) Adjust terminology in the *CSIP* to reflect the language of records creators and systems managers more closely. The Glossary Committee may be able to provide some insight into how to map between the terminology used by interviewees and the terminology being used by the InterPARES Project. Additional analysis of the transcribed tape recordings of the case studies to date should also assist with this aspect.

4) Rewrite the *CSIP* questions so that they use the definitions of the terms, rather than the actual glossary terms used in the Template of Analysis to ask the questions.

5) Eliminate questions that have been demonstrated through three or more rounds of case studies not to yield useful data.

6) Interviewers' comments from the *CSIP* and *TEDGI* suggest that the complexity of the systems being studied need more technical expertise to be fully understood. In a best case scenario, interviews should be conducted with both an archivist and an IT or computer professional.

5. RELATIONSHIP BETWEEN CONCEPTUAL REQUIREMENTS FOR AUTHENTICITY AND EXISTING STANDARDS

5.1 Preamble

In order to place its conclusions in context, the Task Force has conducted comparative analyses of the Authenticity Requirements against three prominent records management standards: the International Standards Organization's (ISO) *Draft International Standard on Records Management*; the U.S. Department of Defense's (DoD) *5015.2 Records Management Standard*; and the European Commission's (EC) *Model Requirements Specification* (MoReq). Each of the "mapping documents" produced in this exercise has been designed to reveal the extent of similarity between the Authenticity Requirements, on one hand, and the particular standard under examination, on the other.[41] The ISO and EC mapping documents identify provisions that can be considered as counterparts to the individual InterPARES Benchmark Requirements, while the DoD mapping locates provisions that function parallel to the stipulations contained in both the benchmark and the baseline requirements.

The mapping documents provide a basis for comparison from both microscopic and birds-eye perspectives. With respect to the former, each mapping reproduces or summarizes individual provisions from the ISO, DoD, or EC standard alongside the particular Authenticity Requirement to which they relate. Thereby, the mapping documents allow for assessment of how InterPARES Requirements are expressed differently from, and similarly to, pertinent provisions of the existing standards. At the same time, the documents can be used to make more general comparisons in that they reveal an overall portrait of the relationship between the InterPARES Requirements and the three existing standards in question. For instance, the mapping documents demonstrate how many of the InterPARES Requirements have counterparts in, respectively, the ISO, DoD, and EC standards.

[41] The mapping documents are available on the InterPARES website.

A brief summary of the findings of each mapping exercise is presented below. Please note that making identifications between provisions of different standards involves recognizing degrees of similarity, and is rarely a simple yes/no question. This is a result of several factors, such as the fact that the specific wording used in any given standard tends to be unique, and the fact that an idea or concept treated in one single provision by, for instance, the Authenticity Requirements might be scattered among several provisions in the ISO, DoD, or EC standard. Therefore, this text generally makes statements to the effect of "DoD provision X is a parallel (or counterpart) to InterPARES requirement Y." Such statements are understood to mean that a general resemblance exists between provisions X and Y, not that they correspond directly and completely with one another. Conversely, the text attempts to avoid suggesting that necessary and complete correspondence is entailed in identification of counterparts and parallels by avoiding statements to the effect of "Fulfillment of DoD provision X satisfies InterPARES requirement Y in all respects."

5.2 International Standards Organization. ISO/DIS 15489: Draft International Standard on Records Management

The ISO Draft Standard is designed to provide "guidance on managing records of originating organizations, public or private, for internal or external clients" by making recommendations "to ensure that adequate records are created, captured and managed." In Section six, the ISO standard indicates that organizations should "establish, document, maintain and promulgate policies, procedures and practices for records management, the objective of which should be the creation and management of authentic, reliable and useable records, capable of supporting business functions and activities for as long as they are required."[42]

There are at least two noteworthy features of ISO/DIS 15489. First, while it provides a considerable amount of technical detail in specifying required software functionalities, the standard also addresses matters like organizational policies and procedures. This establishes an extent of similarity between the ISO standard and the InterPARES Authenticity Requirements in that both guidelines take into account the need for combining automated and manual implementation methods. Second, a particular section of the ISO standard is devoted to emphasizing the importance of record authenticity. This implies further common ground with the InterPARES Requirements, although, in accordance with the orientation of 15489 as a whole, the pertinent sections only treat this matter as it relates to active records.

[42] International Standards Organization, Technical Committee ISO/TC 46 Information and Documentation, Subcommittee 11, Archives/Records Management, *International Standards Organization Draft International Standard (ISO/DIS 15489) Information and Documentation – Records Management* (Geneva: International Standards Organization, 2000), "1. Scope;" in ISO mapping document.

Of the eight Benchmark Requirements, only "A.6 Authentication of Records" and "A.7 Identification of Authoritative Record" were found not to have counterparts within ISO/DIS 15489. On the other hand, parallel provisions from the ISO standard have been identified for each of the remaining six Benchmark Requirements. Note, though, that the counterpart for InterPARES Requirement A.1 concerning "Expression of Record Attributes and Linkage to Record" does not specify any particular metadata fields for capture. Instead, this stipulation indicates that organizations should determine what metadata is required according to their business needs and regulatory circumstances: "To support the continuing conduct of business [and] comply with the regulatory environment … organizations should institute and carry out a comprehensive records management programme which includes … determining what metadata should be created with the record and through records processes and how that metadata will be persistently linked and managed."

In consideration of these parallels, it can be said that full compliance with the 15489—or partial compliance, if all provisions listed in the ISO mapping document were to be satisfied—would result in satisfaction of Requirements A.1 through A.5, and A.8. However, note also that ISO section 7.2.1 on "Authenticity" has been identified as parallel to Requirements A.1 and A.2 only, suggesting that although certain ISO provisions satisfy Requirements A.3, A.4, A.5, and A.8, these would support authenticity only in an implicit fashion.[43]

5.3 United States Department of Defense. Design Criteria Standard for Electronic Records Management Software Applications

The *Design Criteria*, better known as DoD 5015.2, "sets forth mandatory baseline functional requirements, and identifies non-mandatory features deemed desirable" for procurement of records management application (RMA) software by agencies of the United States government. 5015.2 has been implemented in this context for the purpose of "assur[ing] efficient and effective records management." The scope of the standard is restricted to management of active records, and, as a procurement standard, its contents focus almost exclusively on required system functionalities.[44]

---

[43] See ISO mapping document.

[44] United States, Department of Defense, Assistant Secretary of Defense for Command, Control, Communications, and Intelligence Design Criteria, *Standard for Electronic Records Management Software Applications* (DoD 5015.2-STD) June 2001, "C1.1. Purpose." The characterization of the purpose of 5015.2 presented here is based on 44 U.S.C. § 2902, the passage of the United States Code cited in the "C.1.1. Purpose" section of the standard. In full, this law reads as follows:

> It is the purpose of this chapter, and chapters 21, 31, and 33 of this title, to require the establishment of standards and procedures to assure efficient and effective records management. Such records management standards and procedures shall seek to implement the following goals: (1) Accurate and complete documentation of the policies and transactions of the Federal Government; (2) Control of the quantity and quality of records produced by the

These observations suggest some of the ways in which DoD 5015.2 differs from the Authenticity Requirements, and, for that matter, from ISO 15489: specifically, the *Design Criteria* devotes primary attention to software specifications over procedures and other implementation means, and to methods over principles. Furthermore, the focus of DoD 5015.2, like the ISO standard but unlike the InterPARES Requirements, is solely on active records. Finally, this standard is distinct in that it does not overtly address authenticity, the record quality of principal concern to the Task Force, anywhere in its terms.

The DoD standard features provisions that can be understood as counterparts to six of the eight Benchmark Requirements. The exceptions are Requirement "A.6 Authentication of Records," which has no parallel provision in 5015.2, and Requirement A.1 on "Expression of Record Attributes and Linkage to Record," which is satisfied in several respects, although not entirely due to the fact that A.1 mandates capture of certain metadata fields not covered in the DoD standard. As for the InterPARES Baseline Requirements, counterpart provisions have been identified for "B.1 Controls over Records Transfer, Maintenance, and Reproduction," and "B.2 Documentation of Reproduction Process and its Effects." No parallel stipulation was located for "B.3 Archival Description."[45]

5.4 European Commission. Model Requirements for the Management of Electronic Records (Interchange of Data between Administrations (IDA program)

The *Model Requirements Specification*, or MoReq, "focuses mainly on the functional requirements for the management of electronic records by an Electronic Records Management System (ERMS)," and is designed for use by public and private sector organizations that are either introducing an ERMS, or assessing one already in place. MoReq has been designed to be "pragmatic" and "usable," and its purpose is to ensure that an ERMS will "manage electronic records with the desired levels of confidence and integrity."[46]

Federal Government; (3) Establishment and maintenance of mechanisms of control with respect to records creation in order to prevent the creation of unnecessary records and with respect to the effective and economical operations of an agency; (4) Simplification of the activities, systems, and processes of records creation and of records maintenance and use; (5) Judicious preservation and disposal of records; (6) Direction of continuing attention on records from their initial creation to their final disposition, with particular emphasis on the prevention of unnecessary Federal paperwork; (7) Establishment and maintenance of such other systems or techniques as the Administrator or the Archivist considers necessary to carry out the purposes of this chapter, and chapters 21, 31, and 33 of this title.

[45] See U.S. DoD mapping document.

[46] *Requirements for the Management of Electronic Records* (MoReq Specification), prepared by Cornwell Affiliates plc. (CECA-CEE-CEEA: Bruxelles- Luxembourg, 2001), "1.2 Purpose and Scope of this Specification," "1.5 Emphasis and Limitations of this Specification."

Like DoD 5015.2, MoReq is a software specification, and accordingly it differs from the InterPARES Requirements in that it explicitly focuses on system functionality over procedures, and on implementation methods over records management principles. The European Commission standard also shares a point in common with ISO 15489 in that it addresses authenticity of records directly. However, MoReq defines "authenticity" in a manner that may or may not match the InterPARES definition. Note as well that MoReq features a greater extent of variability than any of the other standards considered here, including the InterPARES Requirements. Having been designed to acknowledge that "different countries have their differing traditions, views and regulatory demands for managing records," the EC standard presumes that, prior to use, it will be tailored to the business needs and the legal-regulatory requirements bearing upon an organization.[47]

MoReq counterparts have been located for seven of the eight InterPARES Benchmark Requirements. In the remaining case, several provisions from the EC standard are listed as parallel to Requirement A.1 on "Expression of Record Attributes and Linkage to Record." However, there are metadata fields mandated for capture in InterPARES Requirement A.1 that are not specified in MoReq.[48]

6. RELATIONSHIP OF FINDINGS TO OTHER RESEARCH INITIATIVES

Issues that relate to the preservation and authenticity of digital information objects are being addressed from several perspectives by current research projects. These projects include:

- CAMiLEON (Creative Archiving at Michigan & Leeds: Emulating the Old on the New) is investigating the viability of emulation as a preservation strategy that maintains the intellectual content, structure, and "look and feel" of a software-dependent complex digital objects. Researchers are also assessing user preferences for different versions of emulators that vary considerably in how they reproduce those objects (for example, by analyzing how users define the authenticity of objects running in their native software environment, running under emulation, and delivered in migrated versions.[49]

[47] See "1.5 Emphasis and Limitations of this Specification," "4.5 Authenticity," and "13.1 Glossary." The Glossary defines "authenticity" as "the quality of being genuine," but the quality of genuineness is not itself defined in the MoReq Glossary, or elsewhere. Note also that MoReq does distinguish between "mandatory" and "desirable" requirements, but that organizations implementing MoReq may nevertheless modify mandatory functionalities, and even omit individual requirements, when custom-designing the Specification to suit their business needs.

[48] See EC mapping document.

[49] University of Michigan and University of Leeds. *CAMiLEON: Creative Archiving at Michigan and Leeds. Emulating the Old on the New*. Available: http://www.si.umich.edu/CAMILEON/

- Cornell University's PRISM Project focuses on policy enforcement for ensuring information integrity in the areas of preservation, reliability, interoperability, security, and metadata. PRISM is investigating the long-term survivability of digital information, reliability of information resources and services, interoperability, and security (including the privacy rights of users of information and the intellectual property rights of content creators), and the metadata that makes it possible to ensure information integrity in digital libraries. As part of this project, PRISM researchers carried out one-on-one interviews and discussion groups at Cornell to characterize the current environment and identify digital preservation requirements. They found that few formal policies are in place for distributed resources and that as a result the level of trust about the preservation of content is low. [50]

- The San Diego Supercomputer Center's (SDSC) Collection-Based Persistent Archives and Archivists' Workbench projects are engaged in deriving XML information models from collections of software-dependent data objects and developing tools that can be used to ensure preservation and access to those objects over time. The Persistent Archives approach is built around the OAIS reference model. It supports archival processes from accessioning through preservation and use, and it recognizes the importance of collection-based management. It also exploits inherent hierarchical structures within records, predictable record forms, and dependencies between them. It is designed to be consistent, comprehensive, and independent of infrastructure.[51]

- The Cedars Project (CURL exemplars in digital archives) seeks to address strategic, methodological and practical issues and provide guidance in best practices for digital preservation. Cedars is a United Kingdom collaboration of librarians, archivists, publishers, authors, and institutions (libraries, records offices, and universities). Working with digitized and born-digital materials, Cedars is using a two-track approach to evaluate different preservation strategies through demonstration projects at U.K. test sites; develop recommendations and guidelines; and develop practical, robust, and scaleable models for establishing distributed digital archives. Cedars is also examining other issues related to the management of digital information, including rights management and metadata.[52]

While only one of these research initiatives, the Persistent Archives research at SDSC, focuses specifically on the preservation of electronic records, the Task Force believes that the delineation of the nature of electronic records and the conceptual requirements for authenticity provide a rigorous framework for

[50] Cornell University. *Project PRISM*. Available at: <http://www.library.cornell.edu/preservation/prism.html>.
[51] Rajesekar et al. *Dlib Magazine* (2000).
[52] Cedars Project. *Metadata for Digital Preservation: The Cedars Project Outline Specification Draft for Public Consultation (2000).* Available at: <http://www.leeds.ac.uk/cedars/documents/Metadata/cedars.html>.

approaching issues of preserving the integrity of complex digital objects in general, and electronic records in particular that could be applied in such research initiatives.

7. CONCLUSION

Electronic records are very complex physical objects and intellectual constructs. Both the deductive and the inductive approaches employed by the Authenticity Task Force have constructed a detailed profile of the complexity of contemporary electronic records and identified their embeddedness in their juridical-administrative, provenancial, procedural, documentary, and technological contexts.

In terms of what the Authenticity Task Force learned relating to issues of authenticity, we found that most contemporary records systems are a hybrid of electronic and paper records; that few explicit measures are employed to ensure the authenticity of electronic records and that authenticity is generally assured through procedural means. Authentication technologies only address the authenticity of records over space and in time.

While the Task Force developed a conceptual framework for establishing the requirements for preserving authentic electronic records, it failed to create a single, comprehensive typology of authenticity requirements for electronic records. It identified several possible perspectives from which a typology could be constructed, but none of these can be developed in such a way that they can be thorough, deep, and predictive. It seems likely that a typology based upon individual creators and the acts/procedures/functions they perform would be the single most effective approach. Potentially such typologies could be generalized to other similar settings, but this generalizability would be limited because each creator interprets his or her own juridical context differently and implements it differently procedurally. The Task Force, however, has not at this point collected the necessary data to support such an hypothesis.

In terms of methodological outcomes, the Task Force found that because of the complexity of electronic records and record-keeping, it is both difficult and problematic for those researching or managing electronic records to identify a single, appropriate unit of analysis. Diplomatics approaches the issue from the focus on the individual record, archival science from the perspective of the record aggregate, and systems analysis from that of the automated information or record-keeping system. Each of these perspectives contributes to understanding the nature of the record and its long-term preservation. What is also required, however, is an overall systems approach that takes into account the total record-keeping environment, that is, the sum of all of the contexts identified through InterPARES.

8. AREAS FOR FURTHER RESEARCH

Several areas that need further research emerged out of the work of the Authenticity Task Force. In some cases, these areas amounted to a more sophisticated formulation of questions that initially InterPARES sought to address and found were beyond the scope of a three-year research project. In other cases, new areas emerged in the course of the project. Research questions within these areas can be grouped under three rubrics: theory of the record, technological development, and record-keeping policy:

a. Theory of the record

1. Is it possible to develop an analytical framework that integrates aspects of contemporary diplomatics and archival theory that addresses both the document and record aggregates and identifies and elucidates the role of the different contexts of the records in relation to both the document and record aggregates?

2. Can we provide a more detailed analysis of the various contexts of the records and the ways in which the archival bond might be expressed within those contexts? Can we develop more finely grained instruments that could extract specific aspects of different contexts and tie them to the record in ways that establish the archival bond?

3. Is it possible to develop meaningful typologies of records of specific creators or specific acts, procedures, and functions?

b. Technological development

1. Digital signature technologies have been implemented for the authentication of records across space in time, but what are their implications for the authenticity of electronic records over time? Can digital signatures be extended to assist in the long-term preservation of authentic electronic records, or will their implementation be harmful to the authenticity of the records over time?

2. A related question concerns the infrastructure supporting digital signature technologies. The authority of a digital signature depends on the existence of a public key infrastructure (PKI), which is a hierarchical organization of certification authorities invested with the competence to authenticate the ownership and characteristics of a public key. The effectiveness of such infrastructure depends on the continuity of the chain of trust guaranteed by those certification authorities. As private sector organizations take on the role of certification authorities, what mechanisms are, or should be, in place to guarantee the continuity of the chain of trust in the event that organization ceases to exist?

3. Is it possible to identify different ways in which conceptual authenticity requirements might be addressed or manifested in record-keeping systems in practice? Can these requirements be translated into an implementation context?

c. Record-keeping policy

1. What are the juridical implications of developing a record-keeping system in which some requirements for authenticity are satisfied in an implicit rather than an explicit manner, for example, in a trust management system where the integrity of the system as a whole, including the procedures used to maintain the system creates a presumption of the authenticity of its component parts?

2. To what extent can the models and principles developed by this project for administrative and bureaucratic records be applied to other kinds of digital objects such as records generated for cultural and creative purposes?

InterPARES Project

International Research on Permanent Authentic Records in Electronic Systems

Appendix

REQUIREMENTS FOR ASSESSING AND MAINTAINING THE AUTHENTICITY OF ELECTRONIC RECORDS

Task Force Members:

Heather MacNeil, University of British Columbia (Chair)
Chen Wei, Beijing Municipal Archives
Luciana Duranti, University of British Columbia
Anne Gilliland-Swetland, University of California, Los Angeles
Maria Guercio, University of Urbino
Yvette Hackett, National Archives of Canada
Babak Hamidzadeh, University of British Columbia
Livia Iacovino, Monash University
Brent Lee, University of British Columbia
Sue McKemmish, Monash University
John Roeder, University of British Columbia
Seamus Ross, University of Glasgow
Wai-kwok Wan, Hong Kong Public Record Office
Zhao Zhon Xiu, State Archives of China

28 October 2001

REQUIREMENTS FOR ASSESSING AND MAINTAINING THE AUTHENTICITY OF ELECTRONIC RECORDS

The requirements for assessing and maintaining the authenticity of electronic records that are identified in this document fall into two groups: the first group includes requirements that support the presumption of the authenticity of electronic records before they are transferred to the custody of the preserver[i], while the second group includes requirements that support the production of authentic copies of electronic records that have been transferred to the custody of the preserver. The report is organized into the following sections:

1. Conceptual Framework for the Requirements for Authenticity
2. **Benchmark Requirements Supporting the Presumption of Authenticity of Electronic Records**
3. Baseline Requirements for the Production of Authentic Copies of Electronic Records
4. Commentary on the Benchmark Requirements Supporting the Presumption of Authenticity of Electronic Records
5. Commentary on the Baseline Requirements for the Production of Authentic Copies of Electronic Records

1. Conceptual Framework for the Requirements for Authenticity

1.1 Introduction

Authenticity is defined as "the quality of being authentic, or entitled to acceptance."[ii] *Authentic* means "worthy of acceptance or belief as conforming to or based on fact" and is synonymous with the terms *genuine* and *bona fide*. *Genuine* "implies actual character not counterfeited, imitated, or adulterated [and] connotes definite origin from a source." *Bona fide* "implies good faith and sincerity of intention".[iii] From these definitions it follows that an *authentic record* is a record that is what it purports to be and is free from tampering or corruption.

In both archival theory and jurisprudence, records that that the creator[iv] relies on in the usual and ordinary course of business are presumed authentic. However, digital information technology creates significant risks that electronic records may be altered, either inadvertently or intentionally. Therefore, in the case of records maintained in electronic systems, the presumption of authenticity must be supported by evidence that a record is what it purports to be and has not been modified or corrupted in essential respects. To assess the authenticity of an electronic record, the preserver must be able to establish its *identity* and demonstrate its *integrity*.

The identity of a record refers to the distinguishing character of a record, that is, the attributes of a record that uniquely characterize it and distinguish it from other records. From an archival-diplomatic perspective, such attributes include: the names of the persons concurring in its formation (that is, its author, addressee, writer, and originator); its date(s) of creation (that is, the date it was made, received, and set aside) and its date(s) of transmission; an indication of the action or matter in which it participates; the expression of its archival bond, which links it to other records participating in the same action (for example, a classification code or other unique identifier); as well as an indication of any attachment(s) since an attachment is considered an integral part of a record.[v] The attributes[vi] that establish the identity of a record may be explicitly expressed in an element of the record, in metadata related to the record, or they may be implicit in its various contexts. Those contexts include: its *documentary context*, that is, the archival fonds to which a record belongs, and its internal structure; its *procedural context*, that is, the business process in the course of which the record is created; its *technological context*, that is, the characteristics of the technical components of an electronic computing system in which records are created; its *provenancial context*, that is, the creating body, its mandate, structure, and functions; and its *juridical-administrative* context, that is, the legal and organizational system in which the creating body belongs.

The *integrity* of a record refers to its wholeness and soundness: a record has integrity when it is complete and uncorrupted in all its essential respects. This does not mean that the record must be precisely the same as it was when first created for its integrity to exist and be demonstrated. Even in the paper world, with the passage of time, records are subject to deterioration, alteration and/or loss. In the electronic world, the fragility of the media, the obsolescence of technology and the idiosyncrasies of systems likewise affect the integrity of records. When we refer to an electronic record, we consider it essentially complete and uncorrupted if the message that it is meant to communicate in order to achieve its purpose is unaltered. This implies that its physical integrity, such as the proper number of bit strings, may be compromised, provided that the articulation of the content and any required annotations and elements of documentary form remain the same.[vii] The integrity of a record may be demonstrated by evidence found on the face of the record, in metadata related to the record, or in one or more of its various contexts.

1.2 Benchmark Requirements for Assessing the Authenticity of Electronic Records

The records of the creator belong to one of two categories. The first category comprises those records that exist as created. They are considered authentic because they are the same as they were in their first instantiation. The second category comprises those records that have undergone some change and therefore cannot be said to exist as first created; they are considered authentic

because the creator treats them as such by relying on them for action or reference in the regular conduct of business. However, the authenticity of electronic records is threatened whenever they are transmitted across space (that is, when sent to an addressee or between systems or applications) or time (that is, either when they are in storage, or when the hardware or software used to store, process, or communicate them is updated or replaced). Given that the acts of setting aside an electronic record for future action or reference and of retrieving it inevitably entail moving it across significant technological boundaries (from display to storage subsystems and vice versa), virtually all electronic records belong to the second category. Therefore, the preserver's inference of the authenticity of electronic records must be further supported by evidence – provided in association with the records – that they have been maintained using technologies and administrative procedures that either guarantee their continuing identity and integrity or at least minimize risks of change from the time the records were first set aside to the point at which they are subsequently accessed. The requirements for assessing the authenticity of the creator's electronic records concern this evidence.

1.2.1 The Presumption of Authenticity

A presumption of authenticity is an inference that is drawn from known facts about the manner in which a record has been created and maintained. The evidence that supports the presumption that the record creator created and maintained them authentic are enumerated in the **Benchmark Requirements Supporting the Presumption of Authenticity of Electronic Records** (Requirement Set A). A presumption of authenticity will be based upon the number of requirements that have been met and the degree to which each has been met. The requirements are, therefore, cumulative: the higher the number of satisfied requirements, and the greater the degree to which an individual requirement has been satisfied, the stronger the presumption of authenticity. This is why these requirements are termed 'benchmark' requirements.

1.2.2 The Verification of Authenticity

In any given case, there may be an insufficient basis for a presumption of authenticity, or the presumption may be extremely weak. In such cases, further analysis may be necessary to verify the authenticity of the records. A verification of authenticity is the act or process of establishing a correspondence between known facts about the record and the various contexts in which it has been created and maintained, and the proposed fact of the record's authenticity.[viii] In the verification process, the known facts about the record and its contexts provide the grounds for supporting or refuting the contention that the record is authentic. Unlike the presumption of authenticity, which is established on the basis of the benchmark requirements, this verification involves a detailed examination of the records themselves and reliable information available from other sources about the records and the various contexts in which they have

been created and maintained. Methods of verification include, but are not limited to, a comparison of the records in question with copies that have been preserved elsewhere or with backup tapes; comparison of the records in question with entries in a register of incoming and outgoing records; textual analysis of the record's content; forensic analysis of the medium, script, etc.; a study of audit trails; and the testimony of a trusted third party.

1.3.1 Baseline Requirements Supporting the Production of Authentic Copies of Electronic Records

After the records have been presumed or verified authentic in the appraisal process, and have been transferred from the creator to the preserver, their authenticity needs to be maintained by the preserver. In order to do so, the preserver must carry forward the records in accordance with the baseline requirements that apply to the maintenance of records, producing copies according to procedures that also maintain authenticity.[ix] The production of authentic copies is regulated by the **Baseline Requirements for the Production of Authentic Copies of Electronic Records** (Requirement Set B). Unlike the **Benchmark Requirements**, all of the requirements included in the **Baseline Requirements** must be met before the preserver can attest to the authenticity of the electronic copies in its custody. This is why the requirements for the production of authentic electronic copies are termed 'baseline' requirements.

Satisfaction of these baseline requirements will enable the preserver to certify that copies of electronic records are authentic. Traditionally, the official preserver of the records has been the person entrusted with issuing authentic copies of such records. To fulfill that role, the preserver needed simply to attest that the copy conformed to the record being reproduced. With electronic records, the difficulties related to preservation make it prudent for the preserver to produce and maintain documentation relating to the manner in which it has maintained the records over time as well as the manner in which it has reproduced them to support its attestation of authenticity.

A copy is the result of a reproduction process. A copy can be made from an original or from a copy of either an original or another copy.[x] There are several types of copy. The most reliable copy is a 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.

Any of these types of copy is authentic if attested to be so by the official preserver. By virtue of this attestation, the copy is deemed to conform to the record it reproduces until proof to the contrary is shown. Such attestation is

supported by the preserver's ability to demonstrate that it has satisfied the applicable baseline requirements for maintenance and all of the requirements for the production of authentic copies.

2. Benchmark Requirements Supporting the Presumption of Authenticity of Electronic Records

2.1 Preamble

The benchmark requirements are the conditions that serve as a basis for the preserver's assessment of the authenticity of the creator's electronic records. Satisfaction of these benchmark requirements will enable the preserver to infer a record's authenticity on the basis of the manner in which the records have been created, handled, and maintained by the creator.

Within the benchmark requirements, Requirement A.1 identifies the core information about an electronic record – the immediate context of its creation and the manner in which it has been handled and maintained – that establishes the record's identity and lays a foundation for demonstrating its integrity. Requirements A.2-A.8 identify the kinds of procedural controls over the record's creation, handling, and maintenance that support a presumption of its integrity.

2.2 Benchmark Requirements (Requirement Set A)

To support a presumption of authenticity the preserver must obtain evidence that:

REQUIREMENT A.1: Expression of Record Attributes and Linkage to Record

the value of the following attributes are explicitly expressed and inextricably linked to every record. These attributes can be distinguished into categories, the first concerning the identity of records, and the second concerning the integrity of records.

A.1.a Identity of the record:

- **A.1.a.i** Names of the persons concurring in the formation of the record, that is:
  - name of author[xi]
  - name of writer [xii](if different from the author)
  - name of originator[xiii] (if different from name of author or writer)
  - name of addressee[xiv]
- **A.1.a.ii** Name of action or matter
- **A.1.a.iii** Date(s) of creation and transmission, that is:
  - chronological date[xv]
  - received date[xvi]
  - archival date[xvii]
  - transmission date(s)[xviii]
- **A.1.a.iv** Expression of archival bond[xix] (for example, classification code, file identifier)
- **A.1.a.v** Indication of attachments

A.1.b Integrity of the record:

- **A.1.b.i** Name of handling office[xx]
- **A.1.b.ii** Name of office of primary responsibility[xxi] (if different from handling office)
- **A.1.b.iii** Indication of types of annotations added to the record[xxii]
- **A.1.b.iv** Indication of technical modifications;[xxiii]

REQUIREMENT A.2: Access Privileges

the creator has defined and effectively implemented access privileges concerning the creation, modification, annotation, relocation, and destruction of records;

REQUIREMENT A.3: Protective Procedures: Loss and Corruption of Records

the creator has established and effectively implemented procedures to prevent, discover, and correct loss or corruption of records;

REQUIREMENT A.4: Protective Procedures: Media and Technology	the creator has established and effectively implemented procedures to guarantee the continuing identity and integrity of records against media deterioration and across technological change;

REQUIREMENT A.5: Establishment of Documentary Forms	the creator has established the documentary forms of records associated with each procedure either according to the requirements of the juridical system or those of the creator;

REQUIREMENT A.6: Authentication of Records	if authentication is required by the juridical system or the needs of the organization, the creator has established specific rules regarding which records must be authenticated, by whom, and the means of authentication;

REQUIREMENT A.7: Identification of Authoritative Record	if multiple copies of the same record exist, the creator has established procedures that identify which record is authoritative;

REQUIREMENT A.8: Removal and Transfer of Relevant Documentation	if there is a transition of records from active status to semi-active and inactive status, which involves the removal of records from the electronic system, the creator has established and effectively implemented procedures determining what documentation has to be removed and transferred to the preserver along with the records.

3. Baseline Requirements Supporting the Production of Authentic Copies of Electronic Records

3.1 Preamble

The baseline requirements outline the minimum conditions necessary to enable the preserver to attest to the authenticity of copies of inactive electronic records.

3.2 Baseline Requirements (Requirement Set B)

The preserver should be able to demonstrate that:

REQUIREMENT B.1: Controls over Records Transfer, Maintenance, and Reproduction	the procedures and system(s) used to transfer records to the archival institution or program, maintain them, and reproduce them embody adequate and effective controls to guarantee the records' identity and integrity, and specifically that **B.1.a** Unbroken custody of the records is maintained; **B.1.b** Security and control procedures are implemented and monitored; and **B.1.c** The content of the record remains unchanged after reproduction;

REQUIREMENT B.2: Documentation of Reproduction Process and its Effects	the activity of reproduction has been documented, and that this documentation includes **B.2.a** The date of the records' reproduction and the name of the responsible person; **B.2.b** The relationship between the records acquired from the creator and the copies produced by the preserver; **B.2.c** The impact of the reproduction process on their form, content, accessibility and use; and **B.2.d** In those cases where a copy of a record is known not to fully and faithfully reproduce the elements expressing its identity and integrity, such information has been documented by the preserver, and this documentation is readily accessible to the user;

REQUIREMENT B.3: Archival Description	the archival description of the fonds containing the electronic records includes–in addition to information about the records' juridical-administrative, provenancial, procedural, and documentary contexts–information about changes the electronic records of the creator have undergone since they were first created.

4. Commentary on the Benchmark Requirements Supporting the Presumption of Authenticity of Electronic Records

The assessment of the authenticity of the creator's records takes place as part of the appraisal process. That process and the role of the **Benchmark Requirements** within it are described in more detail in the report of the Appraisal Task Force. This assessment should be verified when the records are transferred to the preserver's custody.

A.1: Expression of Record Attributes and Linkage to Record

The presumption of a record's authenticity is strengthened by knowledge of certain basic facts about it. The attributes identified in this requirement embody those facts. The requirement that the attributes be expressed explicitly and linked inextricably[xxiv] to the record during its life, and carried forward with it over time and space, reflects the Task Force's belief that such expression and linkage provide a strong foundation on which to establish a record's identity and demonstrate its integrity. The case studies undertaken as part of the work of the Task Force revealed very little consistency in the way the attributes that specifically establish the identity of a record are captured and expressed from one electronic system to another. In certain systems, some attributes were explicitly mentioned on the face of the record, in others they could be found in a wide range of metadata linked to the record or they were simply implicit in one or more of the record's contexts. In many cases, certain attributes (for example, the expression of the archival bond) were not captured at all. The Task Force's concern is that, in the absence of a precise and explicit statement of the basic facts concerning a record's identity and integrity, it will be necessary for the preserver to acquire enormous, and otherwise unnecessary, quantities of data and documentation simply to establish those facts.

The link between the record and the attributes listed in Requirement A.1 is viewed by the Task Force as a **conceptual** rather than a **physical** one, and the requirement could be satisfied in different ways, depending on the nature of the electronic system in which the record resides. For example, in electronic records management systems, this requirement is usually met through the creation of a record profile.[xxv] In other types of systems, the requirement could be fulfilled through a topic map. A topic map expresses the characteristics (that is, *topics*) of subjects (for example, records or record attributes) and the relationships between and among them.

When a record is exported from the live system, migrated in a system update, or transferred to the preserver, the attributes should be linked to the record and available to the user. When pulling together the data prior to export, the creator should also ensure that the data captured are the right data. For example, in the case of distribution lists, the creator must ensure that if the recipients specified on 'List A' were changed at some point in the active life of records, the accurate

'List A: Version 1' is exported with the records associated with the first version, and that the second version is sent forward with those records sent to recipients on 'List A: Version 2.'

A.2 Access Privileges

Defining access privileges means assigning responsibility for the creation, modification, annotation, relocation and destruction of records on the basis of competence, which is the authority and capacity to carry out an administrative action. Implementing access privileges means conferring exclusive capability to exercise such responsibility. In electronic systems, access privileges are usually articulated in tables of user profiles. Effective implementation of access privileges involves the monitoring of access through an audit trail that records every interaction that an officer has with each record (with the possible exception of viewing the record). If the access privileges are not embedded within the electronic system but are based on an external security system (such as the exclusive assignment of keys to a location), the effective implementation of access privileges will involve monitoring the security system.

A.3 Protective Procedures: Loss and Corruption of Records

Procedures to protect records against loss or corruption include: prescribing regular backup copies of records and their attributes; maintaining a system backup that includes system programs, operating system files, etc.; maintaining an audit trail of additions and changes to records since the last periodic backup; ensuring that, following any system failure, the backup 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 trail are reflected in the rebuilt files and also guarantee that any incomplete operation is backed up. The capability should be provided to rebuild forward from any backup copy, using the backup copy and all subsequent audit trails.

A.4 Protective Procedures: Media and Technology

Procedures to counteract media fragility and technological obsolescence include: planning upgrades to the organisation's technology base; ensuring the ability to retrieve, access and use stored records when components of the electronic system are changed; refreshing the records by regularly moving them from one storage medium to another; and migrating records from an obsolescent technology to a new technology.

A.5 Establishment of Documentary Forms

The documentary form of a record may be determined in connection to a specific administrative procedure, or in connection to a specific phase(s) within a procedure. The documentary form may be prescribed by workflow control

technology, where each step in an administrative procedure is identified by specific record forms. If a creator customises a specific application, such as an electronic mail application, to carry certain fields the customised form becomes, by default, the required documentary form. It is understood that the creator, either acting on the basis of its own needs, or the requirements of the juridical system, not an individual officer, establishes the required documentary form(s) of records.

When the creator establishes the documentary form in connection to a procedure, or to specific phases of a procedure, it is understood that this includes the determination of the intrinsic and extrinsic elements of form that will allow for the maintenance of the authenticity of the record. Because, generally speaking, that determination will vary from one form of a record to another, and from one creator to another, it is not possible to predetermine or generalise the relevance of specific intrinsic and extrinsic elements of documentary form in relation to authenticity.

A.6 Authentication of Records

In common usage, to authenticate means to prove or serve to prove the authenticity of something. More specifically, the term implies establishing genuineness by adducing legal or official documents or expert opinion. For the purposes of the benchmark requirements, authentication is understood to be a declaration of a record's authenticity at a specific point in time by a juridical person entrusted with the authority to make such declaration. It takes the form of an authoritative statement (which may be in the form of words or symbols) that is added to or inserted in the record attesting that the record is authentic.[xxvi] The requirement may be met by linking the authentication of specific types of records to business procedures and assigning responsibility to a specific office or officer for authentication.

The authentication of copies differs from the validation of the process of reproduction of the digital components of the records. This process occurs every time the records of the creator are moved from one medium to another or migrated from one technology to another.

A.7 Identification of Authoritative Record

An authoritative record is a record that is considered by the creator to be its official record and is usually subject to procedural controls that are not required for other copies. The identification of authoritative records corresponds to the designation of an office of primary responsibility as one of the components of a record retention schedule. The Office of Primary Responsibility is the office given the formal competence for maintaining the authoritative (that is, official) records belonging to a given class within an integrated classification scheme and retention schedule. The purpose of designating an Office of Primary

Responsibility for each class of record is to reduce duplication and to designate accountability for records.

It is understood that in certain circumstances there may be multiple authoritative copies of records, depending on the purpose for which the record is created.

A.8 Removal and Transfer of Relevant Documentation

This requirement implies that the creator needs to carry forward with the removed records all the information that is necessary to establish the identity and demonstrate the integrity of those records. If the system is designed to generate a profile for each record that expresses all the attributes identified in Requirement A.1 it is sufficient to remove the profiles with the records. In the absence of such a profile, it may be necessary to remove and transfer with the records audit trails, indexes, data directories, data dictionaries, and so on.

5. Commentary on the Baseline Requirements for the Production of Authentic Copies of Electronic Records

The establishment and implementation of the baseline requirements take place as part of the function of managing preservation. The preservation function and the role of the **Baseline Requirements** within it are described in more detail in the report of the Preservation Task Force.

B.1 Controls over Records Transfer, Maintenance, and Reproduction

The controls over the transfer of electronic records to archival custody include establishing, implementing, and monitoring procedures for registering the records' transfer; verifying the authority for transfer; examining the records to determine whether they correspond to the records that are designated in the terms and conditions governing their transfer; and accessioning the records.

As part of the transfer process, the assessment of the authenticity of the creator's records, which has taken place as part of the appraisal process, should be verified. This includes verifying that the attributes relating to the records' identity and integrity have been carried forward with them (Requirement A.1), along with any relevant documentation (Requirement A.8).

The controls over the maintenance of electronic records once they have been transferred to archival custody are similar to several of the ones enumerated in the benchmark requirements. For example, the preserver should establish access privileges concerning the access, use, and reproduction of records (Requirement A.2); establish procedures to prevent, discover, and correct loss or corruption of records (Requirement A.3), as well as procedures to guarantee the

continuing identity and integrity of records against media deterioration and across technological change (Requirement A.4). Once established, the privileges and procedures should be effectively implemented and regularly monitored. If authentication of the records is required, the preserver should establish specific rules regarding who is authorized to authenticate them and the means of authentication that will be used (Requirement A.6).

The controls over the reproduction of records include establishing, implementing, and monitoring reproduction procedures that are capable of ensuring that the content of the record is not changed in the course of reproduction.

B.2 Documentation of Reproduction Process and its Effects

Documenting the reproduction process and its effects is an essential means of demonstrating that the reproduction process is transparent (that is, free from pretence or deceit). Such transparency is necessary to the effective fulfillment of the preserver's role as a trusted custodian of the records. Documenting the reproduction process and its effects is also important for the users of records since the history of reproduction is an essential part of the history of the record itself. Documentation of the process and its effects provides users of the records with a critical tool for assessing and interpreting the records.

B.3 Archival Description

Traditionally it has been a function of archival description to authenticate the records and perpetuate their administrative and documentary relationships. With electronic records, this function becomes critical. Once the records no longer exist except as authentic copies, the archival description is the primary source of information about the history of the record, that is, its various reproductions and the changes to the record that have resulted from them. While it is true that the documentation of each reproduction of the record copies[xxvii] may be preserved, the description summarizes the history of all the reproductions, thereby obviating the need to preserve all the documentation for each and every reproduction. In this respect, the description constitutes a collective attestation of the authenticity of the records and their relationships in the context of the fonds to which the records belong. This is different from a certificate of authenticity, which attests to the authenticity of individual records. The importance of this collective attestation is that it authenticates and perpetuates the relationships between and among records within the same fonds.

Endnotes

[i] The preserver is the juridical person whose primary responsibility is the long-term preservation of authentic records. The preserver's responsibilities include appraisal.
[ii] *Oxford English Dictionary*, 2nd ed., s.v. "authenticity".
[iii] *Merriam-Webster Online Dictionary*, s.v. "authentic".
[iv] The creator is the physical or juridical person in whose archival fonds the record exists. The fonds is the whole of the records created (meaning made or received and set aside for action or reference) by a physical or juridical person in the course of carrying out its activities.
[v] An attachment is a document that constitutes an integral part of the whole record, notwithstanding the fact that it exists as a linked, but physically separate, entity.
[vi] The use of the terms *attribute* and *element* in this report should not be confused with the way the terms are used in other contexts, such as the various Standard Generalized Markup Languages (SGML). In this report, a *record attribute* is a defining characteristic of a record or of a record element. A *record element* is a constituent part of the record's documentary form and may be either extrinsic or intrinsic. An attribute may manifest itself in one or more elements of a record's documentary form. For example, the name of the author of a record is an attribute, which may be expressed as a superscription or a signature, both of which are intrinsic elements of documentary form. For a more detailed explanation of the extrinsic and intrinsic elements of documentary form see the Authenticity Task Force's *Template for Analysis* which is available on the InterPARES website. An attribute may also manifest itself in the form of an annotation(s) to a record, in metadata linked to it, or in one or more of its various contexts.
[vii] For example, for an electronic mail message, an authentic copy of a complete message may include only the text. Provided it clearly indicated the author, addressee, receivers, and date as well as the content, it would not need to appear in the same way in which it was seen by the author or addressee. In contrast, an authentic copy of a map would have to retain its original presentation features, including color and feature presentation. Provided these requirements were met, an authentic copy could be produced in GIF, JPEG, or GML format.
[viii] In common usage, *verify* is synonymous with the terms *validate*, *confirm*, *corroborate*, and *substantiate*. According to *Webster's Online Dictionary*, "*validate* means to attest to the truth or validity of something; *confirm* implies the removing of doubts by an authoritative affirmation or by factual proof; *corroborate* suggests the strengthening of something that is already partly established; *substantiate* implies the offering of evidence that sustains the contention."
[ix] It is understood that the records that are maintained by the preserver only exist as copies of the creator's records.
[x] In common language, *copy* and *reproduction* are synonyms. For the purposes of this research, the term reproduction is used to refer to the process of generating a copy, while the term copy is used to refer to the result of such a process, that is, to any entity which resembles and is generated from the records of the creator. An original record is defined as the first, complete record, which is capable of achieving its purposes (that is, it is effective). A record may also take the form of a draft, which is defined as a temporary compilation made for purposes of correction.
[xi] The name of the author is the name of the physical or juridical person having the authority and capacity to issue the record or in whose name or by whose command the record has been issued.
[xii] The name of the writer is the name of the physical or juridical person having the authority and capacity to articulate the content of the record.
[xiii] The name of the originator is the name of the physical or juridical person assigned the electronic address in which the record has been generated and/or sent.
[xiv] The name of the addressee is the name of the physical or juridical person(s) to whom the record is directed or for whom the record is intended.
[xv] The chronological date is the date, and possibly the time, of a record included in the record by the author or the electronic system on the author's behalf in the course of its compilation.
[xvi] The received date is the date, and possibly the time, when a record is received by the addressee.
[xvii] The archival date is the date, and possibly the time, when a record is officially incorporated into the creator's records.
[xviii] The transmission date(s) is the date and time when a record leaves the space in which it was generated.

[xix] The archival bond is the relationship that links each record, incrementally, to the previous and subsequent ones and to all those participate in the same activity. It is originary (that is, it comes into existence when a record is made or received and set aside), necessary (that is, it exists for every record), and determined (that is, it is characterised by the purpose of the record).

[xx] The handling office is the office (or officer) formally competent for carrying out the action to which the record relates or for the matter to which the record pertains.

[xxi] The office of primary responsibility is the office (or officer) given the formal competence for maintaining the authoritative record, that is, the record considered by the creator to be its official record.

[xxii] Annotations are additions made to a record after it has been completed. Therefore, they are not considered elements of the record's documentary form.

[xxiii] Technical modifications are any changes in the digital components of the record as defined by the Preservation Task Force. Such modifications would include any changes in the way any elements of the record are digitally encoded and changes in the methods (software) applied to reproduce the record from the stored digital components; that is, any changes which might raise questions as to whether the reproduced record is the same as it would have been before the technical modification. The indication of modifications might refer to additional documentation external to the record that explains in more detail the nature of those modifications.

[xxiv] For the purposes of this requirement, *inextricable* means incapable of being disentangled or untied, and *link* means a connecting structure.

[xxv] If the attribute values contained in the profile are also expressed independently as entries in a register of all records made or received by the creator, then, in addition to establishing the identity and supporting the inference of the integrity of the record, they would also corroborate such identity and strengthen the inference of integrity.

[xxvi] The meaning of authentication as it is used by the Authenticity Task Force in this report is broader than its meaning in Public Key Infrastructure (PKI) applications. In such applications, authentication is restricted to proving identity and public key ownership over a communication network.

[xxvii] Although, technically, every reproduction of a record that follows its acquisition by the preserver is an authentic copy, it is the only record that exists and, therefore, should normally be referred to as "the record" rather than as "the copy."