



InterPARES 2 Project

International Research on Permanent Authentic Records in Electronic Systems

Diplomatic Analysis

Case Study 19: Preservation and Authentication of Electronic Engineering and Manufacturing Records

Kenneth Hawkins, UBC

February 2006

INTRODUCTION

InterPARES 2 case study 19, *Preservation and Authentication of Electronic Engineering and Manufacturing Records*, examines an engineering experiment testing the authentication of digital model (CAD) records using a content/message/semantic-based methodology rather than media, bit-count, or static provenancial attribute-based authentication. The experiment examined the use of logic and semantics to query a digital entity's meaning to assess its authenticity. The following organizations conducted the activities of the experiment: (a) the originating research partner, which is an agency of the U.S. government with responsibilities in the science, engineering, design and manufacture of complex assemblies, (b) the Research Division of the Electronic Records Archives (ERA) Program, National Archives and Records Administration (NARA) and (c) the San Diego Supercomputer Center (SDSC).

The business context of the test record entities of the experiment is science-based manufacturing of high-assurance, high tolerance machined piece parts and assemblies for the U.S. government. The business owner has an ongoing need to access and use records of these parts and assemblies for business purposes over a long period of time (50+ years) with the assurance that they remain accurate, reliable, and authentic. The records represent complex geometric and topological measurements and relationships of various parts of three-dimensional objects. The abstraction of this information from test record entities in proprietary CAD formats, its expression into enhanced logical forms that support reasoning about part shape and manufacturing actions, rendering into an archival format, sending it across a trusted network and ingesting it into a persistent archive, returning it for verification for authenticity, reliability, and usability form the basis of the study. Therefore, the immediate provenance of the digital entities under study is the bounds of the experiment protocol and the organizational context cited above.

The creation, use and maintenance (including exchange and storage), and disposition of the digital entities in the experiment possess the characteristics of records as defined in the InterPARES glossary, as do the digital entities in use for business purposes by the originating partner.

The purpose of the diplomatic analysis is to assess the status of the identified digital entity as a record. Once the status of the digital entity has been determined, preservation strategies may be proposed by Domain 3.

IDENTIFICATION OF RECORD(S)

A record, as defined by the InterPARES glossary, is a document made or received and set aside in the course of a practical activity. A record must also possess all of the following five components, as established by InterPARES 1 research conclusions: fixed content and form, embedded action, archival bond, persons and contexts. The application of the definition of a record to the creator's digital entities is therefore analyzed according to the following parameters:

1. To be identified as a record, the digital entity must possess fixed content and form,¹ and be affixed to a stable medium (or physical carrier).

In the business activities of the originating research partner, the records set aside include the (1) Pro-Engineer solid model file in CAD native form (binary), (2) the STEP neutral (ASCII) form² and an image of the model in TIFF form. These are stored as an aggregate termed the "bill of materials" in a proprietary product management system. In the scientific activities of the engineering experiment, test record entities of the (2) STEP forms were enhanced into (3) a C++ based horn clause logical form (ASCI) that supported the delineation of additional geometric relationships and reasoning about part shape to create an "authenticating shape fingerprint," (4) Logistica, a proprietary reasoning engine format (binary), and (5) WC3 Ontologic Web Language (OWL) XML format (ASCII), the ultimate archival form. Test record entities in Form (5) were transmitted across a secure extranet between the originating research partner, the Research Division of the ERA Program at NARA, and the SDSC, stored and sent back to the originating research partner, for authentication using the Logistica reasoning engine. Precise specifications of part shapes and relationships were successfully transformed and authenticated using this tool.

2. A record must also participate in an action, defined as the conscious exercise of will by an officer of the creator or by an external person, aimed to create, maintain, modify or extinguish situations. A record results as an unintended by-product or product of the action.

The records set aside in the business activities of the originating research partner, (a) and (b), participate in actions associated with the creation, management, and use of solid model files to design and manufacture machined piece parts and complex assemblies. The test record entities of the engineering experiment, (1) through (5), participate in the actions of the three research partners within the protocol of the experiment.

¹ The InterPARES1 Authenticity Task Force has defined fixed form as the following: 1) binary content of the record, including indicators of documentary form, must be stored in a manner that ensures it remains complete and unaltered, and 2) technology must be maintained and procedures defined and enforced to ensure that the content is presented or rendered with the same documentary form it had when set aside. (See ATF Research Methodology Statement, available at: http://www.interpares.org/documents/interpares_ResearchMethodologyStatement.pdf).

² Standard for the Exchange of Product Model Data (STEP), ISO ISO 10303.

3. A record must possess an archival bond, which is the relationship that links each record to the previous and subsequent record of the same action and, incrementally, to all the records which participate in the same activity. The archival bond is originary (i.e., it comes into existence when a record is made or received and set aside), necessary (i.e., it exists for every record), and determined (i.e., it is characterized by the purpose of the record).

There are at least three aggregates visible among the entities being used here: The relationship of the five (5) entities within the experiment form one type of archival bond; the relationship of elements within each entity that supports delineation and reproduction of its geometric characteristics forms another type of archival bond; finally, entities (1) and (2) represents a single part within a larger assembly of multiple parts known as a bill of material structure which constitutes a third type of archival bond (and is stored in a proprietary product data management system).

4. Record creation must involve at least three persons, whether or not they explicitly appear in the record itself. These persons are the author, addressee and writer; in the electronic environment, one must also take into account two additional necessary persons: the creator and the originator.

- The record's **author** is 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.

For the records set aside in the business activities of the originating research partner, the author is the agency of the U.S. government with responsibilities in the science, engineering, design and manufacture of complex assemblies. For the records of the engineering experiment, the authors are the same agency along with NARA and SDSC.

- The **addressee** the physical or juridical person(s) to whom the record is directed or for whom the record is intended.

For the records set aside in the business activities of the originating research partner, the addressee is the components of the agency responsible for design review, testing, and manufacture. For the records of the engineering experiment the addressees are NARA, SDSC, and, possibly, InterPARES

- The **writer** is the physical or juridical person having the authority and capacity to articulate the content of the record.

For the business activities of the originating research partner, the writers are the design engineers. For the activities of the engineering experiment, the writer is the originating research partner.

- The **creator** is the person in whose fonds the record exists.

For the business activities of the originating research partner, the creators are the design engineers. For the activities of the engineering experiment, the creators are the three research partners.

- The **originator** is the person to whom the Internet account issuing or the server holding the record belongs.

For the records set aside in the business activities of the originating research partner, the originator is the agency of the U.S. government with responsibilities in the science, engineering, design and manufacture of complex assemblies. For the records of the engineering experiment the originators are that same agency along with NARA and SDSC.

5. Finally, a record must possess an identifiable context, defined as the framework in which the action in which the record participates takes place. The types of context include juridical-administrative, provenancial, procedural, documentary, and technological.

- The **juridical-administrative context** is the legal and organizational system in which the creating body belongs.

The engineering experiment is carried out by trusted partners subject to the U.S. laws and regulations governing their agencies and by the provisions of formal Memorandums of Understanding between the partners. The specific legal and organizational context of the originating research partner is subject to legal restrictions and as such is not reportable. The ERA is a program within NARA, an agency governed by the *National Archives and Records Act of 1984*. The ERA is located in College Park, Maryland. The SDSC, at the University of California San Diego, was founded in 1985 through a grant from the National Science Foundation.

- The **provenancial context** refers to the creating body, its mandate, structure and functions.

The originating research partner in the experiment is an element of the U.S. Government with mission responsibilities in the science, engineering, design, and manufacture of complex assemblies. The mandate of the originating research partner subject to legal restrictions and as such is not reportable. The mandate of the ERA Program, part of the U.S. National Archives and Records Administration, is to “authentically preserve and provide access to any kind of electronic record, free from dependency on any specific hardware.”³ The mandate of the SDSC, part of the University of California, is to “support international science and engineering discoveries through advances in computational science and high performance computing”⁴

- The **procedural context** comprises the business procedure in the course of which the record is created.

The procedural context in which records are set aside by persons in the originating research partner’s agency is spelled out in general in the case study final report, especially in section D, narrative answers to the 23 core research questions. Specific procedures, however, are subject to legal restrictions and as such are not reportable.

³ <http://www.archives.gov/era/about/welcome.html>.

⁴ <http://www.sdsc.edu/>.

Diplomatic analysis of **procedural phases** related to the engineering experiment can be broken down as follows:⁵

- a. **Initiative:** the introductory phase of any procedure is “constituted by those acts, written and/or oral, which start the mechanism of the procedure.”⁶

Researchers at the originating research partner created a digital object model of a simple prismatic shape and converted it to an archival format 5) OWL, as described in 1, above. The OWL file was sent to the SDSC’s SRB and stored there for a period, then retrieved by the originating research partner and, after being converted back to format 4), authenticated using the Logistica reasoning tool.

- b. **Inquiry:** this preliminary phase “is constituted by the collection of the elements necessary to evaluate the situation.”⁷

Creators at the three research partners collaborated on the intellectual, scientific, and technical components, protocols, and resources – that is, they designed the experiment.

- c. **Consultation:** this phase is “constituted by the collection of opinions and advice after all the relevant data has been assembled.”⁸

Researchers (writers) at the originating research partner evaluated the experiment results and consulted with researchers (creators) at NARA and SDSC about the same.

- d. **Deliberation:** this phase is “constituted by the final decision-making.”⁹

Researchers (writers) at the originating research partner consulted with NARA and SDSC to pronounce the results of the experiment.

- e. **Deliberation control:** this phase is “constituted by the control exercised by a physical or juridical person different from the author of the document embodying the transaction, on the substance of the deliberations and/or on its forms.”¹⁰

The originating research partner (writer), as the leader of the experiment, performed this role, assisted by NARA and SDSC (creators).

- f. **Execution:** “the documents created in this phase are the originals of those embodying the transactions.”¹¹ In other words, the execution phase results in the issuing of the first record capable of producing the consequences intended by its author.

⁵ The phases of procedure as dictated by Diplomatic Analysis; see Luciana Duranti, *Diplomatics: New Uses for an Old Science* (Lanham, Maryland and London: The Scarecrow Press in association with the Society of American Archivists and the Association of Canadian Archivists, 1998), 115.

⁶ Ibid.

⁷ Ibid.

⁸ Ibid.

⁹ Ibid.

¹⁰ Ibid.

¹¹ Ibid., 116.

The whole activities of the engineering experiment resulted in the setting aside, by each research partner (creator), of the 1) through 5) set of records that were the subject of the experiment.

- The **documentary context** is defined as the archival fonds to which a record belongs and its internal structure.

The relationship of these five (5) entities within the experiment form one type of archival bond; the relationship of elements within each entity that supports delineation and reproduction of its geometric characteristics forms another type of archival bond; finally, entities (1) and (2) represents a single part within a larger assembly of multiple parts known as a bill of material structure which constitutes a third type of archival bond (and is stored in a proprietary product data management system). For a detailed discussion of the complex internal structure of CAD solid object model records, see Appendix A of the case study report.

- **The technological context** is defined as the characteristics of the technological components of an electronic computing system in which records are created.

Apart from the use of a Pro-Engineer CAD system and proprietary product data management system, and the specification of the business-derived file formats (1) and (2) summarized in question 1, above, the technological context of the originating research partner is subject to legal restrictions and, as such, is not reportable. The technological context of the engineering experiment included the SDSC's Storage Resource Broker technology (SRB), (a middleware application that uses grid and metadata technologies to transparently manage data), and the metadata catalogue management systems (MCAT) at each site.¹²

CONCLUSIONS

According to the above analysis, the digital entities comprising the “bill of materials structure,” (1) and (2), as set aside during the business activities of the originating research partner, as well as the digital entities comprising the test records (1) through (5) generated and evaluated during the engineering experiment, have met all the requirements of a record as defined by InterPARES 1.

The authoritative record to be preserved is comprised of both digital and intellectual components. The records set aside during the business activities of the originating research partner comprise digital components and intellectual components that meet its ongoing business requirements and mandate but not its archival requirements. The records of the engineering experiment have expanded digital and intellectual components in that they are the result of scientific activities to employ semantic and logical technologies to authenticate digital records. They have additional structure and content that supports querying the meaning of a digital entity to assess its authenticity.

¹² For more information on SRB, see: <http://www.sdsc.edu/srb/> For more information on MCAT, see: “Digital Archiving and Long Term Preservation: An Early Experience with Grid and Digital Library Technologies,” <http://www.archives.gov/era/papers/thic-04.html>