InterPARES 2 Project
International Research on Permanent Authentic Records in Electronic Systems

Domain 1 Research Questions

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

Geneviève Sheppard, UBC

March 2006

1.1. What types of documents are traditionally made or received and set aside (that is, created) in the course of artistic, scientific, and governmental activities that are expected to be delivered on-line? For what purposes? What types of electronic documents are currently being created to accomplish those same activities? Have the purposes for which these documents are created changed?

- The digital entities pertaining to this case study are born digital as CAD records; the purpose of the case study is to investigate a process designed to guarantee the persistent archivability of these digital records, and to maintain their authenticity over time as these records are translated into different formats, sent across a network and ingested into a persistent archive and then returned for verification of authenticity, reliability and usability.
- The documents created in the course of business activities are:
    - Knowledge-enhanced objects derived from CAD files and STEP files
    - TIFF images of the drawings generated from the CAD model
- The documents created in the course of the archival experiment (which is the case study)
    - Enhanced STEP records that support the description of further geometric relationships and reasoning about the part shape (high tolerance machined piece parts manufactured for the U.S. government)

1.2. What are the nature and the characteristics of the traditional process of document creation in each activity? Have they been altered by the use of digital technology and, if yes, how?

- The processes of document creation in the business context and the context of the archival experiment to examine the authentication of digital model (CAD) records using a content/message/semantic based methodology are as follows:
    - Business context:
        - Product designers create digital entities within proprietary Pro-Engineer

CAD system in the process of manufacturing mechanical piece-part assemblies
        - The proprietary CAD design records are translated into Standard for the Exchange of Product Model Data (STEP) – AP203 format; and a TIFF representation of a two dimensional drawing of the part is created
    - Archival experiment context:
        - The logical form of the STEP record is enhanced into another logical form using C++ based knowledge representation tools
        - These entities are run through a proprietary reasoning engine (Logistica) into WC3 Ontologic Web Language (OWL) XML format
- The original digital entities are altered by the use of digital technology in the interest of preserving these records and ensuring their authenticity, reliability and usability over time

1.3. What are the formal elements and attributes of the documents generated by these processes in both a traditional and a digital environment? What is the function of each element and the significance of each attribute? Specifically, what is the manifestation of authorship in the records of each activity and its implications for the exercise of intellectual property rights and the attribution of responsibilities?

- There are five forms the digital entities take:
    - CAD (binary) – the entity is defined according to geometric shape
    - STEP (ASCII) – the digital entity is organized according to geometric shape
    - Enhanced object form (ASCI)
    - LOGISTICA – proprietary (binary)
    - OWL (ASCII) – syntax defined by W3C
- The digital entities are first generated in the course of business activities and it is the design product engineer who has ultimate responsibility for the geometric solid model created using the CAD system
- Subsequent translations of the digital entities are generated during the engineering experiment activities that prepare the entities for placement in a persistent archive

1.4. Does the definition of a record adopted by InterPARES 1 apply to all or part of the documents generated by these processes? If yes, given the different manifestations of the record's nature in such documents, how do we recognize and demonstrate the necessary components that the definition identifies? If not, is it possible to change the definition maintaining theoretical consistency in the identification of documents as records across the spectrum of human activities? In other words, should we be looking at other factors that make of a document a record than those that diplomatics and archival science have considered so far?

- The records created during the business activities of the originating research partner and the digital entities comprising the test records for the archival experiment do fit the InterPARES 1 definition of a record
- The Diplomatic analysis for this study identified five types of records that possess fixed content and form and are affixed to a stable medium:
    1. Digital solid model CAD files
    2. STEP file and TIFF image of the model
    3. STEP file enhanced into a C++ based horn clause logical form (ASCI)

4. Logistica, a reasoning engine format
5. WC3 Ontologic Web Language (OWL) XML format (ASCII), the final archival form

- 1 and 2 above participate in the action of creating, managing and using the solid model files to design and manufacture machined piece parts; the records pertaining to 1-5 above participate in and are the result of the archival experiment which forms the basis of the case study
- There exists several relationships (archival bond) between the various records:
    - The relationship between the 5 digital entities within the bounds of the archival experiment form one type of archival bond
    - The relationship of elements within each type of digital entity listed above forms a type of archival bond
    - 1 and 2 above are part of a larger assembly of parts that form the bill of material (stored in a product data management system) and this also represents an archival bond
- Three persons (author, addressee and writer) are clearly evident within the regular business processes of the originating research partner
    - U.S. government agency (author, originator), components of agency responsible for design review, testing and manufacture (addressee), design engineers (writer, creator)
- Within the bounds of the archival experiment:
    - Originating research partner, NARA, SDSC (author, addressee, creator, originator)
    - Originating research partner (writer)
- The records also possess an identifiable context:
    - Juridical-administrative: engineering/archival experiment carried out by trusted partners and subject to U.S. laws and regulations governing the agencies and through formal Memorandums of Understanding between the three research partners
    - Provenancial: The three research partners make up the creating body in this case study
    - Procedural: specific phases carried out in the course of the experiment
    - Documentary: refer to explanation of archival bond above
    - Technological: specific to each research partner

1.5. As government and businesses deliver services electronically and enter into transactions based on more dynamic web-based presentations and exchanges of information, are they neglecting to capture adequate documentary evidence of the occurrence of these transactions?

- There is a rigorous change-control process; changes are recorded in the product data management system
- A product data management system captures all digital entities within the scope of creating the digital solid model (entities created in the CAD system)
    - The product data management system captures all actions and transactions that take place within the system

1.6. Is the move to more dynamic and open-ended exchanges of information blurring the responsibilities and altering the legal liabilities of the participants in electronic transactions?

- This question does not appear to apply to the archival experiment that forms the basis of the study (three trusted research partners bound by the Memorandums of Understanding – assume that responsibilities are outlined and understood in terms of the engineering/archival experiment)
- In terms of the business environment in which the digital model (CAD) records are created:
    - The business owner needs to be able to access and use the records for business purposes in its real science-based manufacturing process (the records used for the archival experiment were generated specifically for testing)
    - In the business environment, the design product engineer has ultimate responsibility for the geometric solid model records created using the CAD system

1.7. How do record creators traditionally determine the retention of their records and implement this determination in the context of each activity? How do record retention decisions and practices differ for individual and institutional creators? How has the use of digital technology affected their decisions and practices?

- The business owner must be able to access and use their records for business purposes over a long period of time, over fifty years (retention based on business needs) and,
- Assume there is a legal basis to the retention of records as well since it is a business and is involved in the manufacturing of machined piece parts used by the U.S. government
- Exists concern that the use of technology to preserve the digital solid model records such as encapsulating the CAD file into a STEP file will fail
    - A TIFF image of the drawing is thus also created so that if all else fails, the image of the drawing will survive and a the original model can be reconstructed from the TIFF image
- Problem: there is no way to store construction history of the solid model records in a neutral format. The STEP file only contains that the resultant solid model itself
    - Problematic because it is unlikely that a new model could be constructed from a preserved drawing that would be equivalent, in construction, to the original model
    - To the creators, the construction file is the most important file to preserve, but no way to do so in a neutral format