InterPARES 2 Project

International Research on Permanent Authentic Records in Electronic Systems

Overview

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

Peter Gagné, Université Laval

May 2006

The Creator Context / Activity

Creator: A partnership between an unnamed agency of the US government; the Research Division of the Electronic Records Archives (ERA), which is part of the National Archives and Records Administration (NARA); and the San Diego Supercomputer Center (SDSC).
Creator type: Government focus / Public sphere (central or federal administration)

Juridical context: The unnamed agency is subject to US laws and regulations governing it. The ERA is a program issued under direction of the archivist of the United States. The program is an element of NARA, a US government agency governed by the National Archives and Records Act of 1984. The SDSC is subject to US laws and regulations governing the unit.

In addition, the agencies followed the protocol of the experiment and abided by the provisions of formal memoranda of understanding between one another.

Activity: The activity is the experimentation of techniques to ensure that electronic records (in this case, CAD files) remain accessible by making them self-describing and independent of specific hardware and software (persistent object preservation). In other words, to maintain authentic records over time to enable production of pieces as long as the business requires them and to find solutions to guarantee the persistent archivability of digital records, maintained authentic.

This study conducted an engineering experiment to examine 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. Steps include:

- The use of proprietary Pro-Engineer CAD systems by product designers to create the initial digital entities that aid in the design and manufacturing of mechanical piece-part assemblies.
- The proprietary CAD design records are then translated into Standard for the Exchange of Product Model Data (STEP) AP203 format.

- The logical form of the STEP records is then enhanced into another logical form using C++ based knowledge representation tools.
- These entities are taken through a proprietary reasoning engine (Logistica) into WC3 Ontologic Web Language (OWL) XML format.

This activity can be considered an emerging or nascent business practice, since—as the need for a program such as InterPARES demonstrates—there are as yet no standards, methods or theoretical basis for the long-term preservation of authentic electronic records.

Nature of Partnership

The creator is defined as three separate research partners conducting the engineering experiment in which the creation, use, maintenance and disposition of the digital entities in the case study took place. The research partners in the experiment are:

1) An unnamed agency of the US government with mission responsibilities in the areas of science, engineering, design and manufacture of complex assemblies.
2) The research division of the ERA program, part of the National Archives and Records Administration (NARA).
3) The San Diego Supercomputer Center (SDSC)

The unnamed agency provides the CAD engineering files for use in the experiment. The ERA program and SDSC participation in the experiment centers around archival experiment activities to explore persistent archiving of records in interactive, dynamic and experiential systems. The technical context and infrastructure of the experiment consists of the SDSC Storage Resource Broker and metadata cataloguing system as well as the ERA Virtual Test Lab, which are all linked through a secure government computer network with limited access privileges. There is also an experiment protocol and formal memoranda of understanding between the partners that govern their partnership.

The research division of ERA was established to conduct research jointly with partners within and outside the United States government to assess the feasibility and effectiveness of methods of preserving authentic electronic records. It seeks to *interface with other systems* to provide them with relevant lifecycle management information.

The SDSC, by its community outreach mission, seems to be actively engaged in partnership work, providing the technological material and expertise to parties who need it. Their mission is to extend the reach of scientific accomplishments *by providing* tools such as high-performance hardware technologies, integrative software technologies and deep inter-disciplinary expertise *to the community*.

Bureaucratic/Organizational Structure

<u>Of the Project</u>

There is no direct information on the actual organizational structure of the case study beyond the roles defined for each partner above.

<u>Of the Individual Partners</u>

The originating (unnamed) research partner is part of a US government agency. Its structure of governance is not outlined in the case study final report.

The ERA is a program within NARA, a US government agency and has a program director and executive officer that oversee the management of the program. It is federally funded. The system design of the ERA was awarded through contract to Lockheed Martin Corporation.

The SDSC, a research unit of the University of California at San Diego, is made up of four divisions, four laboratories and two departments. The SDSC has a director, executive director and division director. It is primarily funded by the National Science Foundation. The Center also has a separate director for each division, laboratory and department.

Digital Entities Studied

The digital entities pertaining to this case study are born digital as CAD records. The first research partner is the initial creator of the digital entities serving as the objects of the study and functions to produce CAD solid model files to be used in the design and manufacturing of mechanical piece-part assemblies.

Digital entities for this case study are divided into two categories:

1) Those resulting from business activities
    - Knowledge-enhanced objects derived from CAD files and STEP files
    - TIFF image of the drawing generated from the CAD model
2) Those resulting from the archival experiment:
    - Enhanced STEP record to support the description of further geometric relationships and reasoning about the part shape. These knowledge-enhanced digital object files are the objects of this case study. Created for the purposes of preservation, they cannot be used to help realize mechanical assemblies without first being translated back to STEP and then to the CAD system.

Documentary Practices Observed

<u>Record Creation and Maintenance</u>

Overall, emphasis is more on preservation issues. Creation and maintenance **processes** are not fully discussed in the report. The conceptual or technical connections to other digital or non-digital entities are documented by information schema associated with the product data management system. The company also has a documented business **procedure** defined for the design/manufacture process. "The procedures are documented through a company portal. Each procedure has a process number that can be accessed from a directory. When the process comes up it is shown in a window indicating the steps in the process by sequential numbers. Company employees are encouraged to read the procedures from the portal and not to print them to paper where they become quickly obsolete. There is great concern that the latest procedure be followed." (FR 10) However, it is not known if any of these procedures specifically deal with document creation, description, classification or maintenance.

There are some formal procedures, rules and standards and these are well documented. There is a rigorous **change**-control process. 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.

The Product Data Management System is an operational repository that **stores** the work of product design engineers and records changes. It **captures** all digital entities within the scope of creating the digital solid model (entities created in the CAD system) and transactions that take place within the system.

There appears to be little or no conventions in **file naming**. "Within individual CAD files and the semantic extension formats the representation of each individual attribute or element also has persistent **unique identifiers**. However, the protocol of the engineering experiment did not require the unique identification of each digital entity since there was only one instance of each of the five entities." (FR 8)

The entities are **organized** by the schema of the Product Data Management System. "This schema is developed by configuration management. People are concerned with correct configuration of the part model and drawings that are released to the creator or to external users. The schema does not really reflect the creation process." (FR 13)

It seems that **metadata** is part of the knowledge added to the "knowledge-enhanced" objects. "In this particular case study we develop a way of more quickly understanding the intellectual aspect of the geometry model by attaching as metadata to the solid model its shape feature aspects." (FR 8) When specifically asked about metadata in the final report, the response is that "Metadata are typically name of creator, release version numbers, date of release, etc." (FR 13)

<u>Recordkeeping and Preservation</u>

The experiment itself comprises part of a **preservation strategy** to find a reliable preservation format for the CAD engineering files. Additionally, an interim preservation strategy is identified as the "bill of materials" structure throughout the Final Report. The initial digital entities created by the unnamed agency are transformed/enhanced into STEP records (knowledge-enhanced digital object files) at a later stage for the purposes of persistent archiving/preservation.

However, the creator does not keep the models that it creates in a records management system. In other words, the design engineer who creates the initial CAD model has **no archive** to persistently store his work. The digital solid model is registered in the Product Data Management System, though this system "definitely should not be construed as a persistent archive. By persistent archive we mean an archive that offers the capability to access data, to be ensured that the data has not been tampered with and that the data can still be used in a computer application to support business functions." (FR 3) It seems that this assessment is shared by the users of the Product Data Management system. "The user has no confidence that the PDM system will be persistent. When asked how the engineer himself guarantees that his models will be archived the engineer will report that he has ultimate responsibility for the product model of record not the Product Data Management system. The engineer may actually keep the solid model in his own personal desktop." (FR 4)

With regards to which digital entities the creator considers to be **records**, "The creator considers the generated drawing to be the record of definition, not the model, even though any change to the drawing requires first a change to the model followed by a regeneration of the drawing… The problem may be cultural more than technological. Engineers and craftsmen still prefer to see a drawing spread out [on paper] versus looking at a tiny screen." (FR 12)

The design product engineer does not seem to have the impulse to store the records for **archival reasons**. The perceived need for preservation is instead based on the utilitarian desire to be able to produce or re-produce the models in the future. "The business owner understands that there is a critical, unsolved business requirement to maintain authentic records over time to enable the production of the pieces as long as the business requires them, with the assurance that they meet the same strict standards (tolerances) as the original piece." (FR 4) Archivists appear to be regarded as unnecessary and an encumbrance to the creator. "The people who are the most competent at building solid models and changing them are the people who have access to the models they need to do their day to day business. There is no intermediating (sic) person playing the role of archivist. [Our modelers have] very high access to the model data. They don't have to go to someone else. That would be terribly inefficient." (FR 12)

The generated drawings of the product engineer are registered in the Product Data Management System, an operational repository that stores the work of product design engineers. This system **captures** all the digital entities created in the CAD system. However, the Product Data Management System does not offer the capability to provide access to the data, guarantee the authenticity of the records or ensure their usability in a computer application. The solid model is encapsulated with a STEP file generated from the CAD model as well as a TIFF image of the drawing generated from the CAD model.

With regards to **interoperability** and reduction of specific technological dependence, the mandate of the ERA is to "authentically preserve and provide access to any kind of electronic record, free from dependency on any specific hardware."[1] The basis of this study comprises the abstraction of complex information from proprietary CAD formats into an expression of this information using enhanced logical forms and then rendering it into a "neutral" archival format.

However, there are some concerns with regards to technology change and **obsolescence**. "There is some concern for protection from technology obsolescence by converting CAD files to STEP, but there is no official technical business procedure in place to manage this translation. Creators have concern, as do almost all managers, that translation errors will creep in." (FR 11) There is a 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. As a result, a TIFF image of the drawing is also created so that if all else fails, the image of the drawing will survive and the original model can be reconstructed from the TIFF image. The creator admits that "storing the geometry model in STEP format is not sufficient…we have no migration strategy if the vendor of our Product data management system fails to exist. Currently our high-level product data structure (bills of material) are not saved in standard-neutral format." (FR 15)

What is more, it is a problem that there is no "enabling record"—no way to store information regarding the construction technique of the solid model records in a neutral format. The STEP file only contains the resultant solid model itself. It is unlikely that a new model could be constructed from a preserved drawing that would be equivalent, in construction, to the original

[1] See http://www.archives.gov/era/about/welcome.html.

model. To the creators, the construction file is the most important file to preserve, but there is no way to do so in any **neutral standard form**. "We need to aggressively lobby standards efforts such as STEP to create a standard feature construction history file." (FR 14)

For all this talk of neutral formats and reduction of technological dependence, there is little or no effort to achieve this system-wide. "We have not yet experienced a major technology exchange with our CAD system. It would be a horrific experience if it happened. There has been no **migration** planning or system put in place to assist in migration…There is a very strong desire by the configuration management people to have a technology-neutral product data management system. […] Most of our users are quite aware that vendors are out to trap the users and prevent migration." (FR 13-14)

In this case study, the knowledge-enhanced objects are derived purely for the purpose of **persistent archiving** and not for any other purpose. Once brought back out of the persistent archive, the knowledge-enhanced objects will need to be converted back to STEP format and then to the native CAD file format. "What we are saying is that *there is now more knowledge in the archival form of the solid model than in the operational form*. We believe this will become the rule rather than the exception as more science-based persistent archiving is achieved. In many cases there simply is not enough knowledge in the operational form to guarantee persistent archivability." (FR 4, emphasis in original text)

Accuracy, Authenticity and Reliability

This study includes sending the archival format of the digital entities across a trusted network to form part of a persistent archive and returning it for verification of authenticity, reliability and usability.

<u>Accuracy</u>

Accuracy is not directly addressed in the final report. It is inferred from the text of the report that the notion of accuracy is closely linked to having a digital file that is capable of producing a product that conforms to specifications. To ensure this capability, the creator has the option to assess the geometric quality of the model with a checker. However, "Most creators do not want to take the time to run model quality checks on both sides of the translation [from CAD to STEP]. If problems occur during or after translation, they see any effort to correct these problems as non-productive. When the quality checkers were tried, the creators of the data complained that the quality checker did not give them sufficient data on how to correct the error." (FR 11)

The philosophy of the ERA is archival. It seeks to ensure that electronic records are as accurate decades in the future as they were when first created.

<u>Authenticity</u>

The title of this experiment is sometimes given as "Authenticating Engineering Objects for Preservation." It explores taking the tools used to assess authenticity well beyond the mere listing of static attributes to using logic and semantics to query the digital entity's meaning within a context of manufacturing and business processes.

The creator must maintain these records authentic over time to enable the production of the pieces as long as the **business requires** them. An operational repository exists (Product Data Management System), but it does not offer the capability to provide access to the data, guarantee the authenticity of the records or ensure their usability in a computer application.

Part of the notion of authenticity in this case study is based on the ability to prevent tampering with or **corruption** of the digital entities studied. However, a risk is introduced in the knowledge-enhancement process. "We must not damage the file we are enhancing and we must be able to check the enhanced knowledge file to make sure it has not been corrupted…There is always a risk of knowledge loss whenever a digital format translation takes place. Tools are now becoming available to check solid model STEP files before and after translation, but testing for integrity after a translation is a real cost of archiving and cannot be ignored…This would be absolutely mandatory in a production persistent archive environment." (FR 4-5)

Another part of the notion of authenticity is the ability to identify a part as the one it is said to be, by **identifying** and verifying certain aspects of that part. "In a sense, this case study has worked with methods to discover intellectual aspects of the part that actually assist in the authentication of the part." (FR 8)

The notion of authenticity is also closely linked with that of **provenance** in the eye of the creator. "One can easily imagine that a failure analysis could become part of the provenance of a digital model assembly file. We should not discount the importance of using the full provenance of the digital solid model files (and assembly files) to assist in authenticating the geometry. As a matter of fact, the more precisely the provenance can be assigned to the digital geometry file the better: better to assign the failure of a bearing to the interface between the shaft and the bearing surface than just to assign the failure to the whole assembly with text saying where the failure occurred. Again provenance and authenticity find company." (FR 5)

Assurance of authenticity seems to be equivalent to trusting the system. "Assurance lies in a confidence in the CAD system and its interface to the product data management system." (FR 11)

<u>Reliability</u>

The questions of saving the TIFF image of the model and the lack of a neutral standard format for the preservation of construction information raise questions of reliability for the digital objects preserved in this study. However, "the design/manufacturing engineer does nothing to ensure the reliability and authenticity of the digital entities other than the quality checks." (FR 11)

It appears that the notion of reliability is closely linked to **trust**. This trust takes two forms: (1) trust in the digital entity to do its required job: "Each translation step must be quality assured in some way or the final preservation OWL form will be worthless and not be trusted." (FR 5, footnote 1) and (2) Trust in using the latest version of software to produce correct digital entities. "Support persons ensure that the creator is using the right version of the software." (FR 11)