InterPARES 2 Project

International Research on Permanent Authentic Records in Electronic Systems

Title: **Case Study 19 Final Report: Preservation and Authentication of Electronic Engineering and Manufacturing Records**

Status: Final (public)

Version: 2.0

Submission Date: September 2005, revised September 2007

Release Date: November 2007

Author: The InterPARES 2 Project

Writer(s): Kenneth Hawkins (compiler)
Electronic Records Archives Program, National Archives and Records Administration

Project Unit: Focus 2

URL: http://www.interpares.org/display_file.cfm?doc=ip2_cs19_final_report.pdf

Table of Contents

A. Overview

This case study (hereinafter, CS19) examines, through an engineering experiment, 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 business context of the test record entities of the experiment is science-based manufacturing of high-assurance, high-tolerance machined piece parts for the U.S. Government. The business owner has an ongoing need to access and use these records 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 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.

This engineering experiment builds on InterPARES 1 to examine methods for trusting the content (reliability, accuracy) and authenticity of records as used by the creator and addresses the InterPARES 2 objective of extending research to new record types and aggregations in interactive, dynamic and experiential systems. It explores taking the tools used to assess authenticity well beyond 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 main purpose of the engineering experiment examined by CS19 was to develop an open-source preservation format for digital computer-aided design (CAD) records of solid models used in high-tolerance manufacturing of complex assemblies. The experiment used Web Ontology Language (OWL), a W3C specification that extends XML to allow representation of semantics within metadata schemas, to persist the geometry, topology, and functional characteristics of CAD model objects. The semantic format enabled automated querying of the digital entity's meaning, expressed in its metadata, to assess its authenticity. The intent of the experiment was to preserve not only the geometric specifications of the model but also its semantically encoded metadata, joined to make a "new logical preservation format" for archival purposes. By logical preservation format the experiment partners in CS19 meant a format encompassing not only the fixed form and content of information representing the model but also instructions encoded within its metadata in way that reasoning engines of the future can conduct "proofs" against the object to authenticate it as fit to support the procedural action for which it was designed to be used.

The creation, use and maintenance (including exchange and storage), and disposition of the digital entities in the case study took place within a trusted computational environment consisting of the originating U.S. Government research partner, the Research Division of the Electronic Records Archives (ERA) Program, National Archives and Records Administration (NARA), and the San Diego Supercomputer Center (SDSC). The technical context and infrastructure of the case study experiment includes the SDSC-developed Storage Resource Broker (SRB) and metadata cataloging system (MCAT) and the ERA Virtual Test Lab, all of which are linked through a secure government computer network accessible only to authorized researchers and engineers

.

B. Statement of Methodology

The results of this case study are derived from using the InterPARES 2 methodology for conducting case studies. Written and oral interviews with the originating research partner, senior research and program management staff of ERA, and examination of written reports and presentations on the case study prepared for NARA and international conferences, provided the basis for documenting the engineering experiment in this report. Interim reports were provided to InterPARES 2 researchers and members at the February 2005 plenary and June 2005 international team meetings and comments received at that time have been incorporated into the present report.

The methodology of the engineering experiment itself is conveyed in the narrative answers to the core research questions.

C. Description of Context

Provenancial

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 ERA program of NARA is the U.S. Government's strategic response to the challenges inherent in the diversity, complexity, and enormous volume of electronic records being created by the government today. ERA will be a comprehensive, systematic, and dynamic means for preserving virtually any kind of electronic record, free from dependence on any specific software or hardware used to create it. The SDSC enables international science and engineering discoveries through advances in computational science and high performance computing.

The digital entities in this case study are generated first from business activities and thereafter from translations the entities undergo during engineering experiment activities that prepare the entities for placement into a persistent archive. They originate as model records of discrete machined piece test parts, designed and formulated for testing but in all other respects identical to those that take action in a real science-based manufacturing process. The business processes are subject to legal restrictions and as such are not reportable but, like processes, are used in CAD and CAM environments within private industry and academia.

The immediate provenance of the digital entities under study is the bounds of the experiment protocol and the organizational context cited above.

Juridical-Administrative

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.

Procedural

The original entities (1) are created by product designers using proprietary Pro-Engineer CAD systems and are provided to colleagues charged with computer-aided manufacturing of high-tolerance, high-assurance objects used in complex assemblies. Business rules ensure that the proprietary CAD design record is translated into (2) Standard for the Exchange of Product Model Data (STEP) AP203 format (ISO 10303) and a TIFF representation of a two dimensional "drawing" of the part [this forms an aggregation called a "bill of materials structure," which is maintained by the business owner]. From there the experiment took the logical form of the STEP record and enhanced it into another logical form (3) that supported the delineation of additional geometric relationships and reasoning about part shape using C++ based knowledge representation tools, then taken through a proprietary reasoning engine (4) (Logistica) and into (5) WC3 Ontologic Web Language (OWL) XML format. See responses to questions 1 and 3, below.

Documentary

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). The latter bond appears on the file plan of the approved records schedule of the originating research partner.

Technological

See response to question 4 and *passim*, below.

D. Narrative Answers to the 23 Core Research Questions

1. What activities of the creator have you investigated?

The creator of the digital solid model used in this case study is a mechanical product design engineer. The activities of the creator that this case study has been most interested in are: 1) the act of creating the geometric solid model using the CAD system, 2) the translation of the proprietary CAD geometry model file to a neutral STEP file, and 3) the management of the proprietary CAD geometric solid model file and the management of the STEP neutral file. We have also studied the way in which the designer as a whole manages solid model work products and the meta-data associated with the work products. We determined that the design engineer has no archives in which to persistently store his work. He does have an operational repository (a Product Data Management System) but it 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.

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.

We find that the creator has a very nice and easy-to-use interface from his CAD work station into the Product Data Management (PDM) system but 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.

The following activity is a transformation activity, not an engineering business activity; however, it might in the future become part of a digital archival business activity.

There is an additional activity that we studied that was not performed by the engineer. This was the conversion of the STEP neutral file of the solid model into a (4) knowledge-based form. Another system (5) was used to enhance the STEP file with additional knowledge (e.g., being able to say in the knowledge file that two surfaces are parallel, coplanar or perpendicular to each other). At the time we had not built the link from the engineer's desktop CAD system to the knowledge enhancer system.

2. Which of these activities generate the digital entities that are the objects of your case study?

To answer this question, one must distinguish between business activities and archival experiment activities that prepare the original business data for placement into the prototype persistent archive hosted by NARA and SDSC. Business activities create certain entities which are both proprietary and standard formats, however formats 3-5 are for the purpose of persistent archiving.

The CAD file comes from the business activity to create a solid geometry file. The STEP file allows transportation of the CAD entity to other systems across space. The purpose of the following transformations is to allow transporting the information reliably and accurately across time.

The activity above that enhances the STEP file with additional knowledge also creates a knowledge-enhanced digital object file. The objects in this file are the objects of our case study. Neither the original CAD file nor the STEP file constitute the objects set for our case study but of course the objects of this case study were derived quite formally from the CAD file and the STEP file. It is important to note that in the 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.

Also note that a new 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. Said more technically, the knowledge-enhancement process must make monotonic changes only to the STEP file. 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 case study did not write a solid model checker for the enhanced object file that was actually persistently archived. There simply was not time and money. It is feasible that existing checker software could be converted to check the enhanced object file directly. This would be absolutely mandatory in a production persistent archive environment.[1]

A great deal of effort went into the building the object-oriented and STEP-inspired knowledge-enhancement system. This system was the result of an R&D effort that covered 10 years at a cost of several million dollars. The system was built by engineers and computer scientists with world-class expertise in knowledge-based systems design, machining process planning and STEP. The knowledge enhancement system was created to enable semi-automatic feature-based metal removal. As it turned out we used a lot of the existing code in the knowledge-enhancement system to create the enhanced object file.

3. For what purpose(s) are the digital entities you have examined created?

The digital entities created in the CAD solid model file (from which the enhanced-knowledge objects of interest are derived) are for the physical realization (a fancy phrase for the design and manufacturing) of mechanical piece-part assemblies. Note that, as indicated above, the digital objects of interest (the knowledge-enhanced objects) cannot be used directly to realize mechanical assemblies without a translation back to STEP and then back to the native CAD system. Engineering drawings, process plans, metal removal tool paths and inspection tools paths, just to mention a few, will be derived from the solid model in the process of manufacturing the piece part assembly. [I like to think of the solid model as the "DNA" for manufacturing. It is not wholly accurate but it gets the point across.] The digital solid model may also drive a rapid prototyping tool to create temporary but touchable wax-like 3D physical image of the machine part assembly.

The digital solid model file is also used to investigate assembly failures. 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.

[1] Note to report compiler from engineering experiment partner: The original CAD solid model file went through several translations on the way to the persistent archive format (OWL). First a translation from the CAD proprietary geometry model file to STEP neutral format (Part 21), second from STEP format to an enhanced object form (we call it horn-clause form), third from the enhanced object form to the proprietary format of the Logistica reasoning system and then fourth from the Logistica reasoner into W3C OWL format. I don't want the report to get bogged down in detail, however, each translation step must be quality assured in some way or the final preservation OWL form will be worthless and not be trusted. I have simplified this by just saying we persistently store the enhanced object form. Actually the truth is that we further transform the enhanced-object form eventually into OWL for persistent archival. You can handle this any way you wish to simplify.

4. What form do these digital entities take? (e.g., e-mail, CAD, database)
As indicated earlier the digital objects ultimately take on five different forms:

Form 1: CAD native form (binary)
Form 2: STEP neutral (ASCII)
Form 3: Enhanced object form—horn clause knowledge form (ASCI)
Form 4: Logistica proprietary reasoner form (binary), and finally
Form 5: OWL (ASCII) the ultimate persistent archive form.

Of course the forms we worry most about are the two binary forms, which are proprietary (the native CAD form and the Logistica reasoner form). This forms a suite of archival forms. (see answers to questions 4e, 8, 9 and 18c, and Appendix A, below)

4a. What are the key formal elements, attributes, and behaviour (if any) of the digital entities?
There were five (5) digital entities in the CS19 engineering experiment. The first two entities listed below are produced during actual computer-aided design (CAD) and computer-aided manufacturing (CAM) processes of the original experiment partner. In the actual business process these entities are stored with a TIFF rendition of designs as an archival aggregate in a product data management system. They were extended in CS19's engineering experiment by three additional digital entities. Each iteration of format in the experiment was chosen to either strengthen semantic expressiveness or to capture knowledge representation in a persistent, open source format:

1. The original entities (1) are created by product designers using proprietary Pro-Engineer CAD systems and are provided to colleagues charged with computer-aided manufacturing of high-tolerance, high-assurance objects used in complex assemblies. There is no formal definition of this format in the public domain as the file has a proprietary format.

2. Corporate business rules of the original experiment partner ensure that the proprietary CAD design record (1) is translated into (2) Standard for the Exchange of Product Model Data (STEP) AP203 format (ISO 10303). The formal element, attribute and behaviour definition of the objects in the STEP file are contained in ISO 10303 AP 203. The standard describes the formal representation of the Euler complete boundary representation definition of a solid model. The definition of the elements and attributes are described in an object-oriented representation language called EXPRESS which is ISO 10303 Part 21. EXPRESS schemas are computer-processable and can be verified automatically for syntactical correctness and for the existence of appropriate links to other schemas. Instances of the defined entities form the actual exchanged data. Entity definitions include rules that can be checked at translation time to verify certain aspects of semantic validity of the transferred instances.

3. From there, the experiment took the logical form of this STEP record (2) and enhanced it into another logical form (3) that supported the delineation of additional geometric relationships and reasoning about part shape and action or process

semantics using C++ based knowledge representation tools. The derived features and action semantics able to be represented by this format allow for their automated interrogation by reasoning programs, establishing semantic metadata to enable automated archival authentication of the digital solid model.

4. These entities (3) were then taken through a proprietary reasoning engine (Logistica) to complete rendition of a format (4) with all required attributes and metadata, including the formulation of logical predicates. While the Logistica format is proprietary, it can be said that it contains a knowledge component and a procedural reasoning component.

5. The Logistica entity (4) was converted to Web Ontology Language (OWL) format (5), an open-source, public domain XML specification of the World Wide Web Consortium (W3C) for persistent archiving purposes. The OWL form is in ASCII. The logical components of this form are defined mathematically by concepts of descriptive logic and the syntax of this form is defined by the W3C in the specifications. OWL is a semantic XML format to represent machine interpretable content when the content needs to be processed by applications rather than just structured for presentation to humans. This requirement applies not only to the World Wide Web but to the digital holdings of any given domain within it, including records repositories. OWL can be used to explicitly represent the meaning of terms in vocabularies and the relationships between those terms – in other words, to operationalize an ontology. OWL has more facilities for expressing meaning and semantics than XML, RDF, and RDF-S, and thus OWL goes beyond these languages in its ability to represent machine interpretable content.

4b. What are the digital components of which they consist and their specifications?

There are five digital components in this case study and those are the five different kinds of file types described above. Each file component consists of a solid model whose representation is formally defined in the ISO 10303 STEP.

- Form 1 (CAD form): This form is quite complex and has many components that we do not know of since it is proprietary.
- Form 2 (STEP form): This form is completely defined within ISO 10303.
- Form 3 (Enhanced object form): This form is defined informally and contains a digital subcomponent of geometric relationships not found in ISO 10303.
- Form 4 (Logistica reasoning form): This form's composition is quite complex and is not entirely known to us. It can be basically said that Logistica contains a knowledge component and a procedural reasoning component.
- Form 5 (OWL form): The logical components of this form are defined mathematically by concepts of descriptive logic and are also described by the W3C OWL specification. OWL has many layers of components, which are described in the OWL specification. It is actually quite complex: OWL builds on RDF. RDF builds on XML and XML is the base form. There is almost an object architecture that defines the shape of OWL.

4c. What is the relationship between the intellectual aspects and the technical components?

The intellectual content of a solid model of course is the shape description contained in the interpretation of the solid model file by a CAD system, or in the case of STEP the shape can actually be interpreted by a human. The CAD system is the primary method of converting intellectual aspects of form into technical forms like CAD models, STEP and OWL. 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. The notion of “hole,” “pocket” and “slot” are all intellectual aspects of shape. 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; that is, we can go from technical aspect to intellectual aspect using a reasoning engine. In other words, we have the strong capability in this case study to preserve intellectual aspects of shape. What we are after in this project is intellectual integrity of the part over long periods of time. We use the phrase “message preserving” but could just as well use the phrase “intellectual aspect preserving.” When we use the word “message” we really mean “intellectual message” not a message that says this file contains 10.301 trillion bytes. In some sense it is difficult for us to separate intellectual from technical as they are so intertwined.

Obviously the reason system is agnostic to the intellectual content of the subject matter but certainly not agnostic to the subject matter of logical reasoning. But to a business person logical reasoning is a technical subject that does not concern them.

4d. How are the digital entities identified (e.g., is there a [persistent] unique identifier)?

In the business activities of the originating experiment partner, digital entities (1) and (2), along with a TIFF version of a solid model design, are stored according to documented company policies in a proprietary product data management system (PDM). The PDM in use is a commercial records management application. This aggregate, termed a “bill of materials,” is filed in the PDM according to a numbered schema corresponding with design/manufacturing procedures, and there under by project number. Within digital entities (3), (4) and (5), the underlying format allows the assignment of unique identifiers at the file level depending upon business needs. This is especially true of files formatted according to the ISO 10303 STEP AP203 and part 21 EXPRESS metadata schemas, which among their functions support specification of the bond between components in complex mechanical assemblies. It also should be noted that 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. Furthermore, CS19 is founded on the proposition (already operational in the Semantic Web) that simple enumeration of discrete identity and integrity metadata is inadequate to the demands for discovery and authentication facing the future of archives. The conception of intrinsic documentary form needs to go much further into recognizing the characteristic patterns (classes, relationships, constraints) that cohere among and between otherwise static identity attributes.

4e. In the organization of the digital entities, what kind of aggregation levels exist, if any?[2]

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).

Form 2 STEP form: STEP defines many, many levels of aggregation within the formal definition of the boundary representation of a solid model. Some examples: the geometry of the part consists of aggregates of surfaces, surfaces consist of aggregates of connected curves, curves consist of aggregates of connected lines and lines consist of aggregates of connected points (usually 2). On the topology side: closed topological spaces consist of an aggregate topologically connected faces, faces consist of aggregates of connected edges, and edges consist of connected vertices.

When we go to assemblies we then have aggregates of geometric solid models and in the topology word we have an aggregate of closed (in the Euler sense) topological spaces. In our case study we did not work with assemblies.

In the description of derived features like hole, pocket, cut out, etc., we say that a shape feature is an aggregate of faces but the shape feature does not have to constitute a closed shape. What we are saying here is that the features are aggregates of topological and geometrical elements that are intellectually close to how real humans perceive shape! It is not hard to think of a through hole feature as a cylinder or a pocket as a set of interconnected flat surfaces stitched together.[3]

4f. What determines the way in which the digital entities are organized?

Form 1 (CAD form): Geometric shape defines this organization.
Form 2 (STEP form): Geometric shape defines this organization.
Form 3 (Enhanced object form): “ ”
Form 4 (Logistica reasoning form): “ ”
Form 5 (OWL form): “ ”

5. How are those digital entities created?

Form 1 (CAD form): by CAD system.
Form 2 (STEP form): by translation from CAD form
Form 3 (Enhanced object form): by geometric reasoning system and by translation from STEP form
Form 4 (Logistica reasoning form): by translation from enhanced object form
Form 5 (OWL form): by translation from Logistica reasoning form

[2] When the report compiler asked the originating partner to think about this question in archival terms, a critical examination of archival concepts of aggregation was the result. See Appendix A.

[3] Report compiler note: One of the important transformations that happens in this case study is between the STEP model, which is considered a data form, to a knowledge form. This is crucial step. Without certain key relationships in the form reasoning cannot happen. Data with attributes do not constitute a knowledge form. If one of the attributes represents a relationship to another object then it probably does represent a knowledge form. This is similar to the different between XML Schema forms and RDF forms of XML. RDF is all about connectedness and XML schema is all about forming aggregates of properties to define an object. So the form in which something is stored in an archive has a lot to do with its longevity.

5a. What is the nature of the system(s) with which they are created? (e.g., functionality, software, hardware, peripherals, etc.)

Software.

5b. Does the system manage the complete range of digital entities created in the identified activity or activities for the organization(or part of it) in which they operate?

The business systems (CAD and CAD-to-STEP translators) manage Forms 1 and 2. Archival systems manage Forms 3, 4 and 5.

6. From what precise process(es) or procedure(s), or part thereof, do the digital entities result?

The digital geometric solid models are the result of design/manufacturing process. The objective of the process is to create the shape of the product so that it conforms to specifications.

7. To what other digital or non-digital entities are they connected in either a conceptual or a technical way? Is such connection documented or captured?

The digital solid model is technically connected to many manufacturing digital objects such as engineering release documents, change release notifications, process plans, metal removal tool paths, inspection tool paths, finite element analysis files. The potential number of connections is only limited by the number of relationships installed in the product data management system. These connections are documented by the information schema associated with the product data management system. The number of conceptual connections (implicit/conceptual) is very high including email, design notes, change proposals etc.

8. What are the documentary and technological processes or procedures that the creator follows to identify, retrieve, and access the digital entities?

In most cases the product data management system at the company of interest is integrated into the CAD system. Standard queries are invoked by menu picks from the CAD system. The company has a documented business procedure defined for the design/manufacture process. There is also a capability to access data in the product data management system from a desktop computer. The product data management system is programmed to accepted standard queries. The process is also controlled by a documented company procedure.

9. Are those processes and procedures documented? How? In what form?

Yes, 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. We also have on-line training in the procedures.

10. What measures does the creator take to ensure the quality, reliability and authenticity of the digital entities and their documentation?

The designer follows quality rules regarding the way that they construct the solid model. There is a quality guide for this. Also the creator has the option to assess the geometric quality of the model with a checker. Typically CAD systems have rigorous model checks to make sure that the model construction step performed by the creator does not create bad geometry. The design/manufacturing engineer does nothing to ensure the reliability and authenticity of the digital entities other than the quality checks. Support persons ensure that the creator is using the right version of the software.

11. Does the creator think that the authenticity of his digital entities is assured, and if so, why?

No. They realize that there is no assurance system. There assurance lies in a confidence in the CAD system and its interface to the product data management system. The creators are well aware that the product data management system is not a persistent archive in any sense of the word. 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. Most creators do not want to take the time to run model quality checks on both sides of the translation. 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.

12. How does the creator use the digital entities under examination?

The main job of the creator is to use the model to create a released engineering drawing of the model (complete with tolerances) and to that the drawing is officially registered in the product data management system. The drawing is the definition of record for the solid model, not the solid model. However, when a new drawing must be made the solid model is first changed and then a new drawing is generated. Very few companies in the world use the model as the definition of record.

13. How are changes to the digital entities made and recorded?

The company has a very rigorous change-control process. Changes are recorded in the product data management system.

14. Do external users have access to the digital entities in question? If so, how, and what kind of uses do they make of the entities?

Yes, through the product data management system. External users could be quality inspectors, fabricators, change control agents etc.

15. Are there specific job competencies (or responsibilities) with respect to the creation, maintenance, and/or use of the digital entities? If yes, what are they?

The creators are competent in the use of the CAD system as well as the use of the interface between the CAD system and the product data management system. External users of the data are expected to know the procedures whereby they can use the product data management system to get digital entities. It is never the case that an external user actually retrieves the digital solid model file. External users want paper prints of the drawings or just to view them on their work station. The manufacturing facility is run by drawings not models. The models are just a fast way of generating drawings. Drawing errors are eliminated in orthogonal views since humans no longer do orthogonal projections and angle views etc. This has caused drawing quality basically to shift to model quality.

16. Are the access rights (to objects and/or systems) connected to the job competence of the responsible person? If yes, what are they?

Yes. In our business 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 person playing the role of archivist. We give our modelers very high access to the model data. They don't have to go to someone else. That would be terribly inefficient. See also answer to question 13, above.

The creation, use and maintenance (including exchange and storage), and disposition of the five (5) entities in the case study took place within a trusted computational environment consisting of the originating research partner, the Research Division of the Electronic Records Archives (ERA) Program, National Archives and Records Administration (NARA), and the San Diego Supercomputer Center (SDSC). The technical context and infrastructure of the case study experiment includes the SDSC-developed Storage Resource Broker (SRB) and metadata cataloging system (MCAT) and the ERA Virtual Test Lab, all of which are linked through a secure government computer network accessible only to authorized researchers and engineers.

17. Among its digital entities, which ones does the creator consider to be records and why?

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. There are other digital object files that are generated from the solid model, such as tool path files and inspection path files, which are considered records. The problem may be cultural more than technological. Engineers and craftsmen still prefer to see a drawing spread out versus looking at a tiny screen.

18. Does the creator keep the digital entities that are currently being examined? That is, are these digital entities part of a record keeping system? If so, what are its features?

Yes, the creator treats it as if it were a record because any new drawing must be derived from a new model. The creator does not keep it in a records management system. The model is registered in the product data management system. Again, the creator realizes that the product data management system does not provide persistent archiving capabilities.

18a. Do the recordkeeping system(s) (or processes) routinely capture all digital entities within the scope of the activity it covers?

The digital solid model is not captured by a recordkeeping system. They are captured by a product data management system that can be considered an archive but not a persistent archive. The product data management system does capture all digital entities within the scope of the activity (i.e., all product solid models that are released are kept in the product data management system).

18b. From what applications do the recordkeeping system(s) inherit or capture the digital entities and the related metadata (e.g., e-mail, tracking systems, workflow systems, office systems, databases, etc.)?

In the business activities of the originating experiment partner the digital entities created in the CAD system are captured in the corporate PDM which is a commercial records management application system (cf. question 4d, above). The expression of the experiment digital entities into the final logical preservation format was a process of derivation and extension from both proprietary and open source systems, as detailed in 4a, above. Within the protocol of the CS19 engineering experiment, the digital entities and related metadata were captured by SDSC's Storage Resource Broker and NARA-ERA's Metadata Catalog Management System.

18c. Are the digital entities organized in a way that reflects the creation processes? What is the schema, if any, for organising the digital entities?

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.

18d. Does the recordkeeping system provide ready access to all relevant digital entities and related metadata?

In the business context of the originating experiment partner the PDM system allows ready access to all digital entities and related metadata. Access is accomplished through standard queries invoked by menu picks by such attributes as procedure number, job, creator, design-change number, design release version number, etc. For the CS19 engineering experiment the SRB and MCAT systems provide a variety of means to access digital entities and any combination of metadata. In addition, the experiment protocol called for the logical querying of the semantic metadata of formats (3), (4), and (5), to authenticate the digital entity's identity, integrity and suitability for the manufacturing process for which it was designed.

18e. Does the recordkeeping system document all actions/ transactions that take place in the system re: the digital entities? If so, what are the metadata captured?

The product data management (PDM) system used by the originating research partner in actual business processes captures actions, events and changes to the digital entities (1), (2) and the bill of materials aggregate. Metadata are typically name of creator, release version numbers, date of release, etc. The SRB and MCAT systems captured all changes to the representation of the CAD solid model as it migrated through the semantic format extensions (3), (4) and (5), including the formulation of metadata that support querying by automated reasoning programs.

19. How does the creator maintain its digital entities through technological change?

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. We did have a technology change to our product data management system. This was done manually with few tools and took a very long time to accomplish. There is a very strong desire by the configuration management people to have a technology-neutral product data management system; for instance, one that would be based on a neutral STEP model. I believe the current product data management system has an export facility. We have no way of exporting the construction history of the CAD system to another CAD system. This would not be in the best interest of the CAD vendor to capture the market. Most of our users are quite aware that vendors are out to trap the users and prevent migration.

19a. What preservation strategies and/or methods are implemented and how?

When a solid model is released to the product data management system it 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. The strategy being that if all else fails the drawing will prevail and the model can be reconstructed. Since we have no way of storing the construction history in neutral format, there is no way of preserving the process whereby the creator actually constructed the solid model. The STEP file only contains the resultant solid model. It is highly unlikely that a new model created from scratch from the drawing would be equivalent (in construction technique) to the original solid model. To our creators the construction history file is the most important file to preserve and unfortunately this does not exist in any neutral standard form. Some translators have been constructed by prototype systems but are not product hardened and have not been scaled up to large solid models.

19b. Are these strategies or methods determined by the type of digital entities (in a technical sense) or by other criteria? If the latter, what criteria?

The strategies we have are definitely determined by the type of digital entities.

20. To what extent do policies, procedures, and standards currently control records creation, maintenance, preservation and use in the context of the creator's activity? Do these policies, procedures, and standards need to be modified or augmented?

All of the digital solid model files are controlled by rigid policy for operational use (creation and modification) but none are controlled for preservation with exception of the aggregate mentioned above consisting of solid model, STEP model and .TIFF drawing image. Yes, I think these procedures need to be augmented. We need to aggressively lobby standards efforts such as STEP to create a standard feature construction history file.

21. What legal, moral (e.g., control over artistic expression) or ethical obligations, concerns or issues exist regarding the creation, maintenance, preservation and use of the records in the context of the creator's activity?

Our products have long life cycles and we are obligated to ensure the customer that these products can be maintained over long periods of time. We must be able to weather technology obsolescence and changes in vendor status. The creator should be able to access the solid model from the product data management system and be able to make changes to it many years hence without having to completely reconstruct the solid model or re-tolerancing the model on a new CAD system. We do not today have the equivalent longevity in the digital solid model that we have in the velum drawing. We did have an experience about fifteen years ago in which a computer-aided drafting vendor filed for bankruptcy and the electronic legacy drawings had to be reconstructed in the new CAD system. This was before the day of the solid model and the generation of drawings from the solid model. This reconstruction cost many millions of dollars. Without the ability to store solid model construction history and tolerances in neutral format we are faced with the same situation if our current vendor becomes bankrupt. Again storing the geometry model in STEP format is not sufficient. Also 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. I do believe we have a way to export them to a flat file form, but of course this flat-file form reflects the vendor's data structures and would most likely have to be translated into another form for a new vendor.

22. What descriptive or other metadata schema or standards are currently being used in the creation, maintenance, use and preservation of the recordkeeping system or environment being studied?

CS19's first digital entity (1), produced during actual computer-aided design (CAD) and computer-aided manufacturing (CAM) processes of the original experiment partner, originates in a proprietary software tool; thus, the precise metadata schema is unavailable. However, the tool produces models in conformance with the ANSI Y-14.5 tolerance standard and provides export files (2) compliant with ISO 10303 Standard for the Exchange of Product Model Data (STEP), AP 203 and part 21 EXPRESS. Corporate metadata standards and procedures govern the filing of these two digital entities with a third TIFF export of the model view into a commercial Product Data Management System. The formats of CS19's digital entities (3) and (4) included the formulation of additional semantic metadata by in-house computer scientists expert in knowledge representation systems that supported the delineation of additional geometric relationships of the CAD solid model and reasoning about part shape and action or process semantics. While some of the metadata supporting action semantics was lost in the translation to digital entity (5), OWL XML, it was able to persist and authenticate precise specifications about part shapes and relationships, including the classes, predicates and constraint rules that govern the identity and behaviour of the CAD solid models.

23. What is the source of these descriptive or other metadata schema or standards (institutional convention, professional body, international standard, individual practice, etc.?)

ANSI, ISO, W3C and corporate business rules. See also answer to question 22.

E. Narrative Answers to Applicable Domain and Cross-Domain Questions

"Domain 3 Template Questions," September 6, 2005

1. What types of entities does the diplomatic analysis identify in this case study? (i.e., records, publications, data, etc.)

The entities in this case study are generated first from business activities and thereafter from translations the entities undergo during engineering experiment activities that prepare the entities for placement into a persistent archive. They originate as model records of discrete machined piece test parts, designed and formulated for testing but in all other respects identical to those that take action in a real science-based manufacturing process. The original entities (1) are created by product designers using proprietary Pro-Engineer CAD systems and are provided to colleagues charged with computer-aided manufacturing of high-tolerance, high-assurance objects used in complex assemblies. Business rules ensure that the proprietary CAD design record is translated into (2) Standard for the Exchange of Product Model Data (STEP) AP203 format (ISO 10303). From there the experiment took the logical form of this STEP record and enhanced it into another logical form (3) that supported the delineation of additional geometric relationships and reasoning about part shape using C++ based knowledge representation tools, then taken through a proprietary reasoning engine (4) (Logistica) and into (5) WC3 Ontologic Web Language (OWL) XML format. 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) represent 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).

1a. Are the entities reliable? If not, why not?

The creation, use and maintenance (including exchange and storage), and disposition of the five (5) entities in the case study took place within a trusted computational environment consisting of the originating research partner, the Research Division of the Electronic Records Archives (ERA) Program, National Archives and Records Administration (NARA), and the San Diego Supercomputer Center (SDSC). The technical context and infrastructure of the case study experiment includes the SDSC-developed Storage Resource Broker (SRB) and metadata cataloging system (MCAT) and the ERA Virtual Test Lab, all of which are linked through a secure government computer network accessible only to authorized researchers and engineers.

1b. Are the entities accurate? If not, why not?

The business rules of the originating partner require that the CAD digital models and drawings produced by them meet the ANSI Y-14.5 tolerance standard, which means that measurements can be made and assured down to the millionth of an inch. The purpose of the experiment is to enhance these proprietary digital entities and their static identifying attributes with enhanced semantic knowledge about the geometric and topologic characteristics of the model and to associate this knowledge with an archival form of the model permanently. As such these characteristics may support a level of accuracy for preservation and reproduction far exceeding authentication based on verification of static attributes. However the full range of knowledge representation added to the entities did not translate into the archival format, so within the context of the engineering experiment they may be said to not be completely accurate.

2. For what purposes are the entities to be preserved?

The originating partner has a need to be able to use its records to satisfy their original purpose over long periods of time and for the records to support the manufacture of machined piece parts to exact specifications or else the failure of parts and assemblies becomes a significant risk. The preservation of the follow-on entities address purposes of the business process and more immediately of the engineering experiment (as detailed in question 1, above).

3. To what degree can the entities be presumed to be authentic, and why?

By the exacting standards of the engineering experiment the entities cannot be said to be authentic to a satisfactory degree. The business-derived entities (1) and (2) form two-thirds of a bill of materials structure that is generated and managed by competent authorities within a trusted, secure environment. Static attributes establish the identity and integrity of the entities considered to be records by the business owners. However the attributes and metadata applied to these entities that may allow permanent storage of rich authenticating knowledge about their geometric and topologic characteristics and relationships did not survive or translate through all the entities of the experiment. The knowledge-enhanced entities (3) and (4, produced by the proprietary reasoning engine) outpaced the ability of the final archival form (5) to preserve this advanced information.

4. Given our knowledge of current preservation methods, can the entities be preserved?

The results of the engineering experiment show that full preservation of the digital entities, including the static attributes of (1) and the rich knowledge representation captured in (4) may only be preserved as long as the proprietary companies stay in business and the originating partner maintains license agreements with them. At present it is not possible to preserve all elements and components of the entities in a persistent archives.

5. If yes, what elements and components need to be preserved?

The originating partner and research partners maintain that to fully preserve and authenticate the digital entities to meet business requirements for long-term archiving, reduce risk and protect investment, preservation must extend considerably beyond the bill of materials *fond* currently being stored in product data management system.

6. Which preservation method might most usefully be applied, and what are its strengths and weaknesses? Which alternative preservation methods might be applied, and what are their strengths and weaknesses?

For the time being the *status quo* method of preserving flat ASCII files of the bill of materials structure through storage in a proprietary product data management system is the most useful means of preservation available to the originating partner. The partner recognizes the acute shortcomings of this approach. Having to leave substantial investments of knowledge and budget in formats subject to proprietary controls outside of the enterprise can and has introduced risks that are costly to overcome to meet archival business requirements. The alternative methods of preservation under examination in the engineering experiment have the advantage of beginning to capture and preserve rich semantic knowledge *about* the objects and processes used to manufacture them; they move toward providing not only the traditional provenancial *what* answers that support identity requirements but also the *why* answers that help support authentication and preservation at an advanced level.

7. What metadata is required to support appraisal and preservation? If metadata is missing, where should it come from and how should it be made manifest?

The standard attributes supporting identity and integrity can and are applied to the original entities now. The attributes that are missing in the archival format (5) are those that support additional reasoning about the characteristics and relationships of part shape to become a form of knowledge representation of the object. Some of these attributes and metadata are "trapped" in the proprietary format (4) but also shortcomings in the OWL specification do not allow for their representation there.

8. Are there any policy constraints that would affect the preservation of the entities?

The main policy constraints that affect the preservation of the entities are market-driven in that the tools that allow additional knowledge representation about the digital entity to be generated are proprietary software programs whose underlying code is not in the public domain. Additionally, there are mathematical constraints to OWL (5) that make it intractable to support the action semantics necessary to preserve the knowledge generated in format (4).

F. Bibliography

Anonymous, "Authenticating Engineering Objects for Digital Preservation," November 16, 2004, College Park, MD.

Gladwell, Malcom, *Blink: The Power of Thinking Without Thinking* (Boston, 2005).

Hawkins, Kenneth, "Status Report on an InterPARES 2 Case Study, CS 19," InterPARES Workshop 13, February 21, 2005, Vancouver, BC.

Hawkins, Kenneth, "'Set-Aside' in the Semantic Web: Findings and Implications of Other Government Case Studies," paper presented at L'Associazione Nazionale Archivistica Italia – ANAI Seminario Internazionale "I risultati di InterPARES2," Centro Congresso Le Stelline, Milan, Italy, 12-13 December 2006.

U.S. InterPARES Project, "Findings on the Preservation of Authentic Electronic Records, Final Report to the National Historical Publications and Records Commission (Grants #99-073 and #2001-005)," International Research on Permanent Authentic Records in Electronic Systems, September 2002.

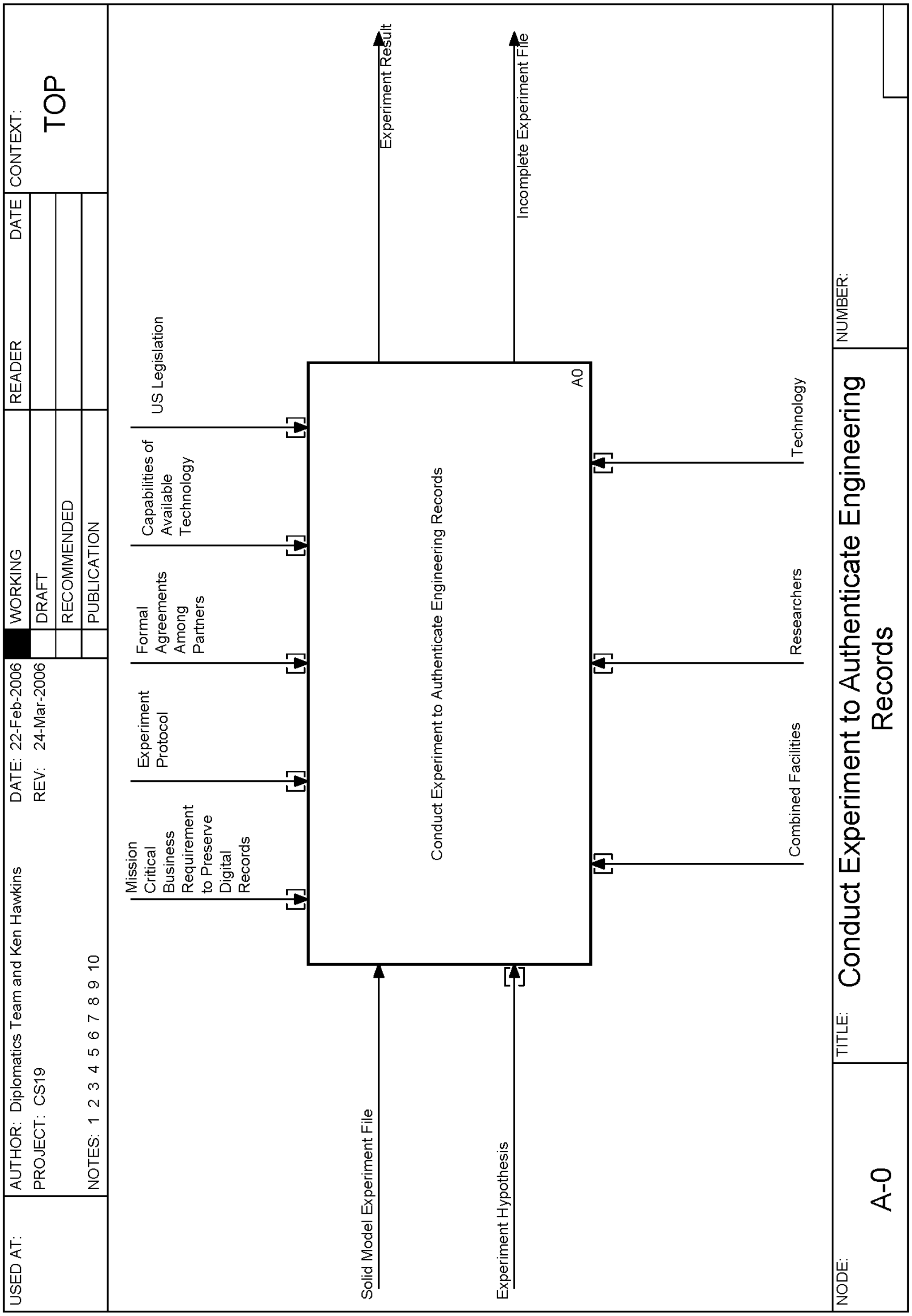
G. Glossary of Terms

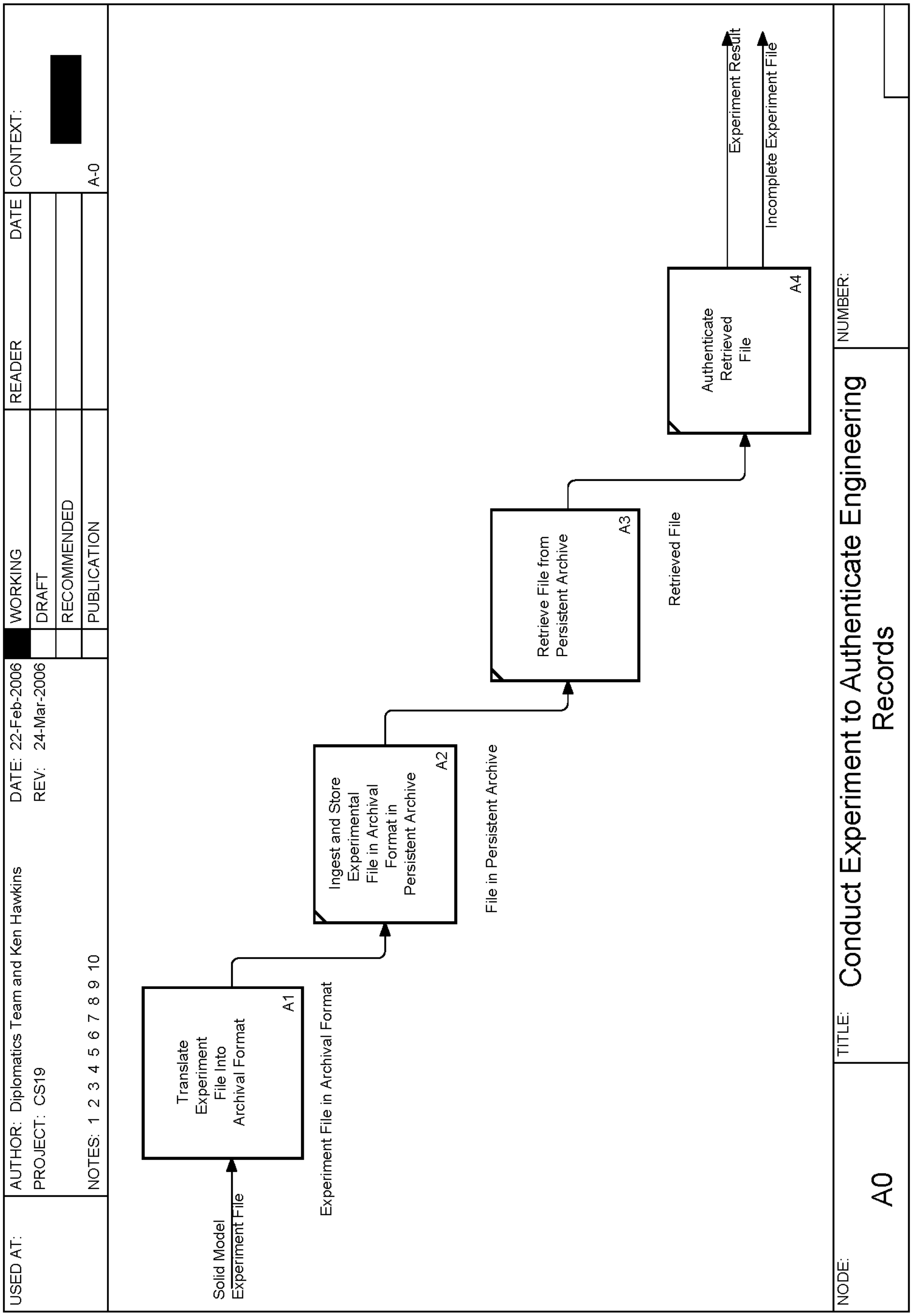
Bill of Materials Structure: the tripartite documentary fonds created by the originating research partner business owners of proprietary CAD file, STEP file and TIFF drawing.

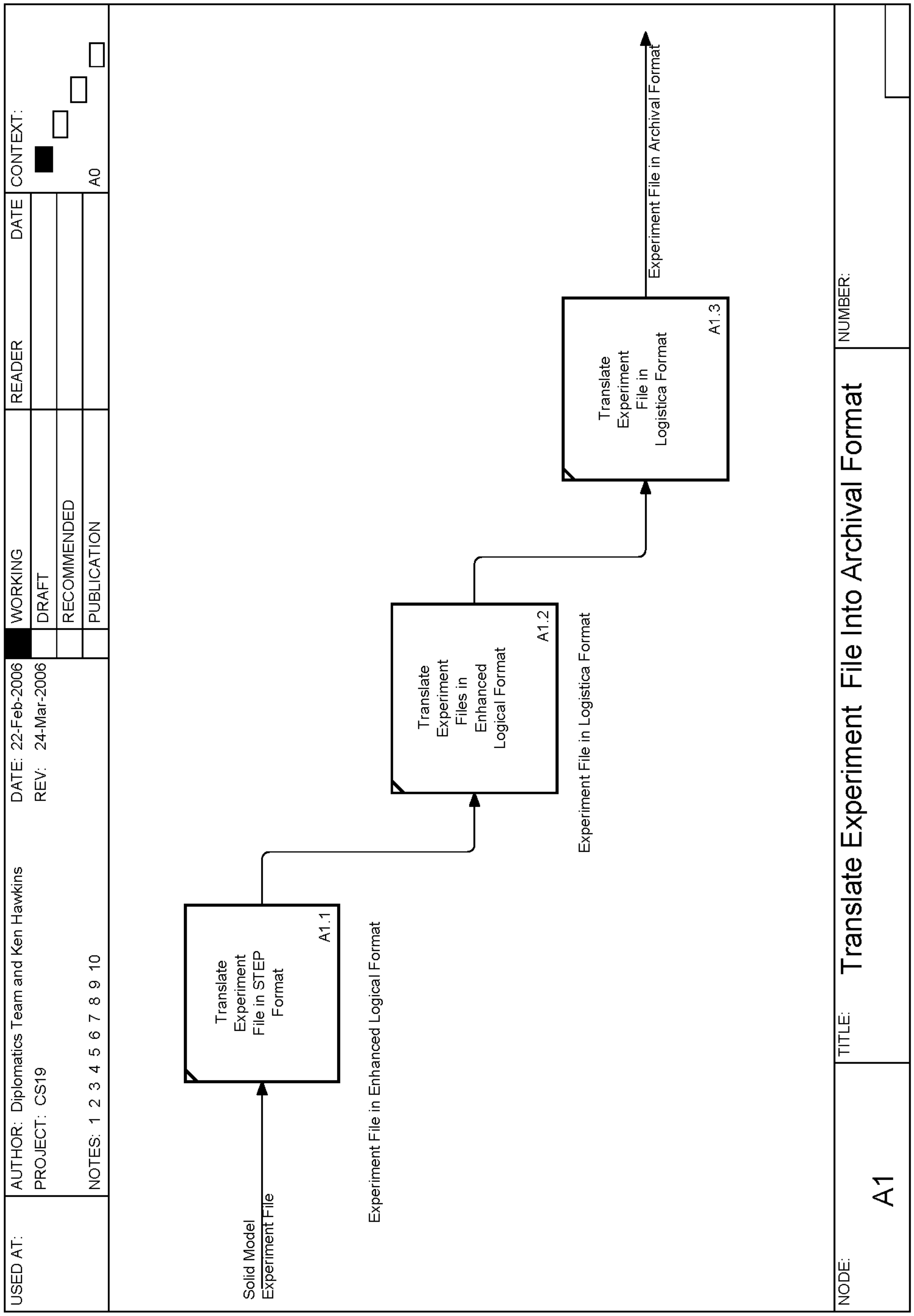
CAD: Computer-Aided Design

CAM: Computer-Aided Manufacturing

H. IDEF0 Activity Model




USED AT:
AUTHOR: Diplomatics Team and Ken Hawkins
PROJECT: CS19
NOTES: 1 2 3 4 5 6 7 8 9 10
DATE: 22-Feb-2006
REV: 24-Mar-2006
WORKING
DRAFT
RECOMMENDED
PUBLICATION
READER
DATE
CONTEXT:
A-0
Solid Model Experiment File
Translate Experiment File Into Archival Format
A1
Experiment File in Archival Format
Ingest and Store Experimental File in Archival Format in Persistent Archive
A2
File in Persistent Archive
Retrieve File from Persistent Archive
A3
Retrieved File
Authenticate Retrieved File
A4
Experiment Result
Incomplete Experiment File
NODE:
A0
TITLE:
Conduct Experiment to Authenticate Engineering Records
NUMBER:


USED AT:
AUTHOR: Diplomatics Team and Ken Hawkins
PROJECT: CS19
NOTES: 1 2 3 4 5 6 7 8 9 10
DATE: 22-Feb-2006
REV: 24-Mar-2006
WORKING
DRAFT
RECOMMENDED
PUBLICATION
READER
DATE
CONTEXT:
A0
Solid Model Experiment File
Translate Experiment File in STEP Format
A1.1
Experiment File in Enhanced Logical Format
Translate Experiment Files in Enhanced Logical Format
A1.2
Experiment File in Logistica Format
Translate Experiment File in Logistica Format
A1.3
Experiment File in Archival Format
NODE:
A1
TITLE:
Translate Experiment File Into Archival Format
NUMBER:

CS19 – Preservation and Authentication of Electronic Engineering and Manufacturing Records, IDEF0 Model Activity Definitions (20060316)			
Activity Name	**Activity No.**	**Activity Definition**	**Activity Note**
Conduct Experiment to Authenticate Engineering Records	A0	To conduct an experiment to authenticate engineering objects in digital formats for long-term preservation.	This experiment was conducted by the originating experiment partner (whose name is withheld for privacy reasons), the ERA program of NARA, and San Diego Super Computer Center.
Translate Experiment File Into Archival Format	A1	To translate experiment file in STEP format; to translate experiment file in enhanced logical format; and to translate experiment file in Logistica format.	The archival format of the experiment files that was the output of activities in A1 is Web Ontology Language (OWL), an open-source, public domain semantic XML format. OWL was formally recommended by WC3 in 2004 for use as a language to represent machine-interpretable content "when the information contained in documents needs to be processed by applications, as opposed to situations where the content only needs to be presented to humans. OWL can be used to explicitly represent the meaning of terms in vocabularies and the relationships between those terms."
Translate Experiment File in STEP Format	A1.1	To transform STEP format into a horn-clause logical form using C++ that supports delineation of additional geometric relationships, reasoning, and knowledge representation about part shape.	The topological characteristics represented in the experiment file output by the activities in A1.1 related to convexity of surfaces and surface area.
Translate Experiment Files in Enhanced Logical Format	A1.2	To transform experiment files into the Logistica reasoning format that supports further reasoning about the solid model.	Logistica is a proprietary format.
Translate Experiment File in Logistica Format	A1.3	To transform experiment file from the Logistica format to Ontologic Web Language (OWL) XML format in order to render the enhanced experiment file and associated metadata into a persistent archival format.	See Activity Note A1, above.
Ingest and Store Experimental File in Archival Format in Persistent Archive	A2	Experiment file in archival format is sent across a trusted network to SDSC for storage in ERA prototype.	
Retrieve File from Persistent Archive	A3	To locate and retrieve experiment file stored in archival format from the persistent archive for authentication.	
Authenticate Retrieved File	A4	To verify the extent to which knowledge representing the geometric and topological characteristics and relationships in the file in the enhanced Logistica format was transferred into the archival format.	To do this, OWL files are rendered back into the Logistica format and assessed for the completeness of the enhanced metadata originally generated in the earlier steps of the experiment.

CS19 – Preservation and Authentication of Electronic Engineering and Manufacturing Records, IDEF0 Model Arrow Definitions (20060316)

Arrow Name	Arrow Definition	Arrow Note
Capabilities of Available Technology		
Combined Facilities	Physical plant and networks shared by research workers.	
Experiment File in Archival Format		
Experiment File in Enhanced Logical Format	Experiment file translated into enhanced logical format.	Logistica is a proprietary format that allowed the addition of ‘action semantics” about the experiment object which the reasoning engine in Logistica could apply to deduce or authenticate features of the object.
Experiment File in Logistica Format	Experiment file translated into Logistica format.	
Experiment Hypothesis	Proposal that authenticity of complex electronic records can be ensured accurately through the use of logic- and semantic-based procedures and tools.	
Experiment Protocol	Rules, procedures, and controls that regulate scientific experiment, including geometry and formal logic.	Procedures include quality assurance checks at each step of the experiment.
Experiment Result		
File in Persistent Archive	The experimental file in archival format that has been stored and ingested in the persistent archive.	
Formal Agreements Among Partners	Memorandums of Understanding between experiment partners, as well as contractual agreements between NARA and SDSC, and the accepted InterPARES2 case study proposal.	
Incomplete Experiment File		
Mission Critical Business Requirement to Preserve Digital Records	Business requirement to maintain authentic records over time to enable production of machine piece parts and complex assemblies for as long as the originating research partner requires them.	
Researchers	Scientists, engineers, programmers and archival experts of the three research partners.	
Retrieved File	Experiment file in archival format retrieved from persistent archive.	
Solid Model Experiment File		
Technology	Hardware and software resources that allow the experiment to take place.	See CS19 Final Report for details about the technological context of the experiment.
US Legislation	US laws and regulations governing US agency research partners, as well as those controlling information and business activity security.	

Appendix 1: Digital Aggregates

From: <@>
To: “Kenneth Hawkins” <ken.hawkins@nara.gov>
Date: 8/22/05 12:38PM
Subject: Re: Last questions answered

Ken,

You are touching on a subject that is near and dear to my heart - aggregations!

I was under the assumption that InterPARES II was had a much broader and more abstract interpretation of the “physical” fonds. Yes, I spent many days studying the notion of fonds in the last couple of years. I am very familiar with the structure of STEP files as physical objects. I have created solid models from scratch by creating physical STEP files and then sending the STEP files to solid modeling systems. I did this to test my understanding of the conceptual schemas (Express models) in STEP - particularly the Euler geometry model. I feel that the physical relationships of the actual entities in the STEP file (each line in the file has a unique identifier just like a record would have in a fonds) is a kind of fonds. However there is a logical fonds that organizes one thinking about the physical fonds of the STEP file. I have spent hundreds of hours validating the physical STEP files and if it were not for having the logical/conceptual model (the Express Schema) I would have had difficulty “navigating” through the physical fonds of the STEP file. I think in a sense a records management system has a layout of boxes and indices within boxes that forms a kind of logical fonds that creates a roadmap to the physical fonds (the boxes). So this idea is not really new but it has been “miniaturized” and applied to the physical breakdown of a single file of a single solid model part. It would be no different than having an XML DTD for a document in which the document was broken down into sections, chapters, paragraphs and sentences. The DTD acts as a road map for the physical fonds, which is the organization of the elements of the digital object within the XML document.

Based on the InterPARES I principles of preservation, the first principle is that digital objects have complex structure and can be constituted by subordinate digital objects, which implies that the fonds has complex structure that goes beyond simple physical contiguity. I made that immediate assumption as soon as I read that principle. I can't imagine having any other interpretation. So it nearly frightens me to hear that the fonds still being referred to in the former way by the InterPARES research group.

I say all of this to say that I have a very broad notion of the concept of a fonds, which goes far beyond physical contiguity in a paper sense. I abstract it to mean the possible physical relationships of lines in a file - I carry it one step further. I got this idea two years ago when I was doing my study on fonds. I don't have the paper at my finger tips but in summary the paper (I think it was written by an Aussie) suggested the notion of a logical or conceptual fonds or intellectual fonds etc. The notion of connectivity in a complex object must of course go beyond sequential physical location in a box of paper.

I think in terms of mathematical lambda function (the lambda calculus) in which almost any set of objects can be created based on an open formula (read query) against a consistent set of highly related data for which there is a schema or roadmap. A simple query such as “please find all of the surfaces on part X523 and organize them by surface area and print the results on a piece of paper” is just as much a physical fonds as the original physical file but I have strayed off the path.

Let me get back:

What I am saying is that within a physical file for a solid model that has 500 surfaces, there will be 500 individual lines in the file that are the root entry for a surface. I count the collection of these lines (even though they may be scattered throughout the physical file) as the surface fonds of the total part fonds. Now if a given surface has a set of say six very complex spline curves which as a connected aggregate define the boundary of that single surface I count that collection as the boundary fonds for the surface fonds. Again the boundary elements could be scattered throughout the physical file. Now each boundary element in the boundary fonds consists of an aggregate of line segments that constitute the boundary element. Some times a single line segment, if it is a very complex curve and depending on the accuracy of the geometry (it could be a millionth of an inch), can be a megabyte in size!

It is hard for me to understand how InterPARES can move the work forward without a broad interpretation of fonds. I myself simply could not do it.

I hope this is helpful and please call me as this is a crucial issue in complex object authentication in general not just for geometry. When we discovered the features of the solid model in our case study we had surface fonds flying all over the place and we literally had hundreds of lambda functions being generated dynamically in order to find certain sub-fonds of interest that would lead our non-monotonic reasoner to the conclusion that we had actually discovered a geometric feature.

So our entire research effort is founded on the idea of dynamic fonds and logical/conceptual fonds described with descriptive logics.

I don't know what to do except tell the people in your report that we just changed our thinking about fonds in order to get the job done!