

The International Journal of Digital Curation

Issue 2, Volume 3 | 2008

Sustaining Engineering Informatics: Toward Methods and Metrics for Digital Curation

Joshua Lubell, Sudarsan Rachuri, Mahesh Mani,
National Institute of Standards and Technology, USA

Eswaran Subrahmanian,
Carnegie Mellon University

July 2008

Abstract

Ensuring the long-term usability of engineering informatics (EI) artifacts is a challenge, particularly for products with longer lifecycles than the computing hardware and software used for their design and manufacture. Addressing this challenge requires characterizing the nature of EI, defining metrics for EI sustainability, and developing methods for long-term EI curation. In this paper we highlight various issues related to long-term archival of EI and describe the work towards methods and metrics for sustaining EI. We propose an approach to enhance the Open Archival Information System (OAIS) functional model to incorporate EI sustainability criteria, Digital Object Prototypes (DOPs), and end user access requirements. We discuss the end user's requirements from the point of view of *reference*, *reuse* and *rationale* – the “3Rs” – to better understand the level of granularity and abstractions required in the definition of engineering digital objects. Finally we present a proposed case study and experiment.

Introduction

Engineering informatics (EI) is the study of the design and use of information structures and systems that facilitate the practice of engineering to achieve socio-economic goals. It is a discipline supporting codification (syntax and semantics), organization, exchange, sharing, decision-making, storage, and retrieval of digital objects characterizing the multi-disciplinary domain of engineering. This is a difficult problem, as it requires combining a diverse set of emerging theories and technologies, e.g., information science, product engineering, and various engineering specialties (Subrahmanian & Rachuri, [2008](#)).

Even without addressing issues specific to engineering, the general problem of long-term digital preservation is complex and open-ended. Issues include:

- Limited understanding of long-term archiving requirements
- Lack of a cost/benefit model to rationalize archiving
- Lack of formal methods and standards tailored to specific application domains for long-term retention of knowledge
- Inefficient archival procedures
- Lack of clear policy guidelines
- Lack of clear metrics and archival methodology
- Limited institutional support for archiving.

Additionally, it is critical for engineers that the digital models and systems they build today be extensible and reusable by subsequent generations of technologists. But even though many products have lifecycles spanning multiple decades (e.g., aircraft, ships, power generation equipment), design repositories and product lifecycle management systems assume that data are always readable (Kopena, Shaffer & Regli, [2006](#)). This assumption is questionable at best when a digital product model has a longer lifespan than the data formats, application software and computing platforms used to create the model. And data must be writable as well as readable if a digital product model, or its supporting information, needs editing at some point during the product's lifecycle.

To address these challenges, we aim to characterize the nature of EI, develop methods for sustaining long-term usability of EI artifacts and define metrics for digital curation¹. Our immediate focus is on archiving engineering information from Computer-Aided Design (CAD), Computer-Aided Engineering, Computer-Aided Manufacturing, Product Lifecycle Management and related software applications used in creating engineering documents. These digital documents are of many different formats, both proprietary and standards-based. This underscores the complex nature of the problem of digital format sustainability.

We outline a technical approach and roadmap which, if followed, could be a significant contribution to the digital archiving of EI. It is expected that these results could lead to standards for creating engineering-oriented digital object prototypes (Saidis & Delis, [2007](#)) –domain-specific realizations of digital object structures – and

¹ The Digital Curation Centre <http://www.dcc.ac.uk> defines *digital curation* as “maintaining and adding value to a trusted body of digital information for current and future use; specifically, the active management and appraisal of data over the life-cycle of scholarly and scientific materials.”

the process for creating and using these archives. The methodology and metrics could also be applied to other scientific disciplines, such as, chemistry, biology, and other areas where critical information must be “future-proofed.”

Our approach is based on an ongoing study of end user access requirements and existing models of archiving in the literature, including standards efforts and other methods deployed in digital libraries and in governmental archives. As part of this activity, we are creating a set of long-term sustainability criteria and a classification system for digital objects relevant to engineering. This effort will serve as the basis for evaluating the archiving of the different kinds of digital objects present in the engineering community.

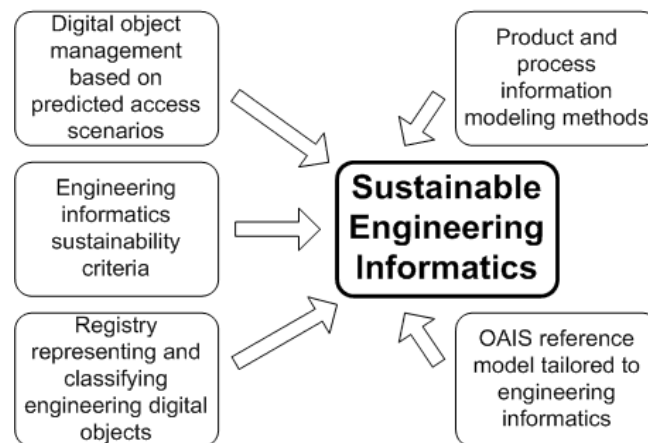


Figure 1. Overview of EI sustainability approach.

Figure 1 summarizes key inputs. To achieve EI long-term sustainability, we consider the following ingredients to be necessary:

- Representation methods for both product and process information
- A strategy based on anticipation of future access requirements for managing archived digital objects
- Sustainability criteria and metrics tailored for EI
- A registry representing and classifying EI digital objects
- Extensions to the Open Archival Information System (OAIS) reference model (Consultative Committee for Space Data Systems [CCSDS], [2005](#)) to address issues specific to EI.

The rest of this paper is organized as follows. The next section provides a summary of related work. We then present our ongoing efforts to develop sustainability metrics and classification criteria. Lastly we suggest next steps toward achieving long-term sustainability for EI and a proposed experiment.

Related Work

There are five main strands of related work. Our ongoing activities and roadmap for the future aim to synthesize these related efforts to create a customized framework for the EI domains. Evaluation of the developed framework can then be accomplished through a prototype implementation and test bed for creating an EI archive. In the sciences and engineering, digital representations often have an underlying information model. It is thus important to determine sustainability factors for such structured

digital objects. Taken a step further, it would be of great benefit to establish sustainability metrics for other aspects of an archival system besides the digital formats used, e.g., to be able to directly measure the quality of an ingest process or an access mechanism.

The first strand of work by Robert Kahn of the Corporation for National Research Initiatives and Robert Wilensky of the University of California at Berkeley (2006) addresses the problem of distributed digital object storage and the infrastructure facilities required for its management. They define a model akin to the Domain Name Server (DNS) model for internet address resolution through the use of a Uniform Resource Identifier (URI) which they call a digital object identifier (DOI). Their model goes beyond just indexing the object for retrieval, since it also has other facilities such as a retrieval access protocol that manages the retrieval process using the DOI to retrieve the objects from the distributed store. In this model, objects may be replicated for reasons of efficiency of retrieval. Their model also includes features for creating mutable and immutable objects allowing for modification of the digital object content without changing the URI when the object context is in flux. The digital object can be rendered immutable for permanent and unalterable digital object storage. When an object is immutable it can be replicated across the network without fear of future inconsistency. The Handle System^{2 3}, a service providing unique, resolvable identifiers for Internet objects, is an implementation of Kahn and Wilensky's research results.

The second strand of work for indexing and managing digital objects comes from projects such as Pergamos⁴ from the University of Athens, Carnegie Mellon University's Typed Object Model (TOM) (Wing & Ockerbloom, 2000), and Fedora (Lagoze, Payette, Shin & Wilper, 2005). TOM and Pergamos employ type-object taxonomies to allow for the identification of the given digital object format, the possible translations among these different formats, and the ability to retrieve the digital object based on the end-user tool that is available for displaying the content. Fedora uses semantic web technologies to represent distributed information and to associate digital objects with web services. These projects, which build on the digital object framework of Kahn and Wilensky, form the infrastructural pieces for constructing and managing a digital archive but are generic architectures that are not specific to the EI world.

Pergamos creates a software environment for storing digital objects using the idea of a digital object prototype (DOP). Pergamos uses the digital object type model but does not enforce a strict hierarchy of types. Types are treated as prototypes and hence modifiable to suit end users' needs. Once a prototype is defined, there is strict adherence to the prototype specification in the definition of the digital object instances. There is a model that allows for the composition of the digital objects. Pergamos is based on a flexible information model for creating DOPs and instances, thereby allowing for a structured organization of the digital objects themselves. Like Fedora, Pergamos is compatible with the creation of a distributed object store that is critical for engineering and science informatics.

² The Handle System <http://www.handle.net>

³ Mention of commercial or third-party products or services in this paper does not imply approval or endorsement by NIST nor does it imply that such products or services are necessarily the best available for the purpose.

⁴ Libraries Computer Center, University of Athens, Pergamos Digital Library <http://pergamos.lib.uoa.gr/>

The third strand of work comes from information modeling of product and process information in EI. In the EI world there is a necessity to manage different data structures, formats, and compositions that are particular to engineering. For example, an assembly of a collection of parts, a configuration of a product and a configuration of information, for say, tolerance or assembly analysis, are specific information models for specific tasks in engineering. To represent EI, we need product and process information models that cover the entire lifecycle of the product. The National Institute of Standards and Technology's (NIST) Core Product Model (Fenves, 2001), Open Assembly model (Sudarsan et al, 2004), and other work in part-part relations and compositions, as well as the work of the Methodology and Tools Oriented to Knowledge-based Applications (MOKA) project,⁵ can form the basis for the information models that can be used to guide the creation of appropriate DOPs corresponding to the different levels of abstraction in the information model. DOPs provide a domain-specific realization of digital object types and classes in the context of digital libraries. The range of information exchange between participants in a design has also led NIST to the creation of a typology of standards according to the information model and content exchanged, based on the expressivity of the information model.

The fourth strand corresponds to the work on the reference model for the process of archiving digital objects. The OAIS Reference Model (CCSDS, 2005) is a generic model for archiving that is achieving wide acceptance in a variety of domains. An engineering-specific application of OAIS is LOTAR (LOng Term ARchiving)⁶, a proposed International Organization for Standardization (ISO) standard for creating archival information packages using the Standard for the Exchange of Product Model Data (ISO 10303) – informally known as STEP (the STandard for the Exchange of Product model data) (Kemmerer, 1999). Commercial products based on LOTAR are in development in the aerospace industry. One such effort⁷ uses STEP to represent geometry and product data management information.

The fifth strand of work involves standards and tools for managing the multitude of digital formats in an archival system. XML-based methods for packaging digital objects include METS (the Metadata Encoding Transmission Standard)⁸ and the PREMIS (PREservation Metadata: Implementation Strategies) data dictionary⁹. Software tools for identification, validation, and characterization of digital objects include JHOVE (JSTOR¹⁰/Harvard Object Validation Environment)¹¹ and DROID (Digital Record Object Identification)¹². Several projects are developing format

⁵ MOKA: A Framework for structuring and representing engineering knowledge
<http://web1.eng.coventry.ac.uk/moka/>

⁶ ProSTEP iViP Association – LOng Term Archiving – LOTAR
<http://www.prostep.org/en/project-groups/long-term-archiving-lotar.html>

⁷ EPM Technology EXPRESSway <http://www.epmtech.jotne.com/newsletter/ew106/eurofighter.html>

⁸ Metadata Encoding and Transmission Standard (METS) Official Web Site
<http://www.loc.gov/standards/mets/>

⁹ PREMIS : Preservation Metadata Maintenance Activity (Library of Congress)
<http://www.loc.gov/standards/premis/>

¹⁰ JSTOR is an acronym meaning “Journal Storage”. Thus, JHOVE is an acronym containing another acronym!

¹¹ JSTOR/Harvard Object Validation Environment <http://hul.harvard.edu/jhove/>

¹² DROID (Digital Object Record Identification) <http://droid.sourceforge.net/>

registries using these tools^{13 14 15}. The Library of Congress's digital format registry¹⁶ is particularly relevant to our work because it provides sustainability criteria, as discussed in the next section (see *Sustainability and Classification of EI Digital Objects*).

The related work presented above provides the strands of research that are critical to the realization of a framework for archival of digital objects in engineering and science. The challenge in EI is to integrate these strands in order to create a truly distributed, user-managed (without requiring computer programming) and context-sensitive model for archiving digital objects.

Present Efforts: Toward Sustainability Metrics

The ability to replicate the behavior of the artifact or the experiment in the validation of science and engineering knowledge is crucial. This requires that the information be available in the best form for retrieval and reuse. The need to know a designer's intent becomes important in the context of redesign and reuse of existing parts. Another important aspect of engineering archiving is the ability to store the digital objects at different levels of granularity and abstractions as required by the design decision-making tasks. Without such an ability to compose different digital objects for archiving it would not be possible to maintain the ability to encode reuse or rationale-based access needs.

We therefore consider end-user needs from the point of view of *reference*, *reuse* and *rationale* – the “3Rs” – to better understand the level of granularity and abstractions required in the definition of digital objects. By “end user” we mean what the OAIS reference model refers to as the *designated community*.

We are also classifying the digital standards that are used in the design and manufacture, as well as throughout the lifecycle of the product, using a previously developed typology (Subramanian, Sudarsan, Fenves, Fofou, & Sriram, 2005). This effort will yield results regarding what standards are used, where they are used, and how are they expected to be used from the end-user perspective. This work can also be used to define the DOPs that are specific to the standards. We also focus on critical issues like original quality of digital formats in archiving, information loss due to transformation from one digital format to another, and importance of standards-based formats for archiving. Our goal is to define better sustainability metrics and a framework for well-defined policies of digital curation.

In a nutshell, our approach amounts to enhancement of the OAIS functional model to include pre-processing based on EI sustainability criteria plus DOPs and post-processing based upon the 3Rs as shown in Figure 2. The post-processing portion of the figure illustrates the hierarchy of access requirements for the 3Rs. The capabilities needed for reusing information are a superset of those sufficient for merely referencing the information. The access requirements for understanding rationale subsume those for reuse.

¹³ PRONOM <http://www.nationalarchives.gov.uk/pronom/>

¹⁴ Global Digital Format Registry <http://www.formatregistry.org/>

¹⁵ FOCUS (Format Curation Service) <http://www.umiacs.umd.edu/research/adapt/focus/>

¹⁶ Sustainability of Digital Formats: Planning for Library of Congress Collections <http://www.digitalpreservation.gov/formats/>

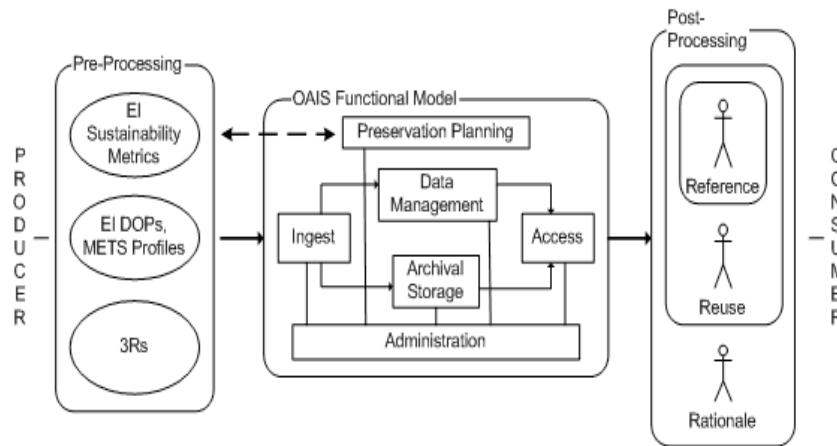


Figure 2. Enhanced OAIS functional model.

The 3Rs: EI Designated Community Requirements

The 3Rs – reference, reuse, and rationale – define a taxonomy of designated community access scenarios. By *reference* we mean the ability to read the digital object and produce the digital object for proper reproduction in a given display medium (computer display, paper, etc.). We use the term *reuse* to mean the ability to refer to and modify the digital object in an appropriate system environment (software and hardware). The *rationale* is the highest level of access in which the end user should be able to refer, reuse and explain the decisions about the content of the digital object.

The primary driver for the 3Rs is the special retrieval needs for each of these scenarios. For example a collection intended primarily for reference may need to be organized differently than one intended for reuse, where not only the geometric aspects of the product are sought but also additional information regarding manufacturing, part performance, assembly and other aspects. In a similar vein, rationale information may have to be packaged differently in that it may include requirements information along with other performance data on the part or the assembly. Given the range of uses and perspectives of the end users, their needs will have a large impact on the process of archiving and retrieval.

Figure 3 illustrates these terms using as an example a STEP representation of a gear object. STEP physical files use an ASCII format defined in ISO 10303-21 (2002). The STEP processor can be any software application capable of interpreting and/or generating STEP physical files, for example a CAD tool capable of importing and exporting STEP files, or a visualization tool that can import STEP data.

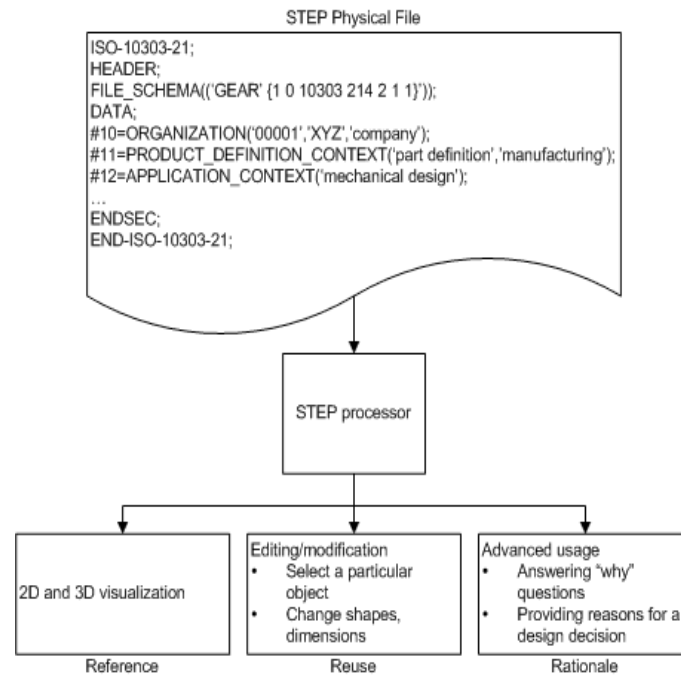


Figure 3. STEP and the 3Rs.

Toward an Archival information Metrology for EI

A standardized EI digital object within a specified set of conventions has a form (syntax), function (scope) and the ability to convey as unambiguously as possible an interpretation (semantics) when exchanged. The design of a standardized digital object in the context of information metrology is dictated by the following parameters:

1. **Language:** the symbols, conventions and rules for encoding and expressing content. Examples include first order logic, OWL (Web Ontology Language)¹⁷, and UML (Unified Modeling Language)¹⁸.
2. **Processable Expressiveness:** the degree to which a language mechanism supports machine understanding or semantic interpretation. Expressiveness is closely connected to the scope of the content that can be represented and to the precision associated with that content. Support of standardized exchange requires a set of complementary and interoperable standards.
3. **Content:** the information to be communicated. Content includes the domain's information model, the information model's instances, and an explanation of the relationship between the message and the behavior it intends to elicit from the recipient. Examples of content include STEP (Kemmerer, 1999), NIST Core Product Model (CPM) (Fenves, 2001) and its extensions (Sudarsan et al, 2004).
4. **Interface:** *User interface* concerns efficiency of communication between the system and humans. *Software interface* concerns accuracy and completeness of communication between systems.

By analyzing the 3Rs using the above four parameters, we can build formal definitions of reference, reuse and rationale and determine the end-user requirements for EI. Engineering digital objects need to be shared in a collaborative and secure

¹⁷ Web Ontology Language (OWL) <http://www.w3.org/2004/OWL/>

¹⁸ Object Management Group – UML <http://www.uml.org/>

manner across the global enterprise and its extended value chain. It is absolutely critical that the sharing mechanism preserve semantic correctness and be efficient, inexpensive, and secure. In order to create such a sharing mechanism, consistent standards, measurements, and specifications are needed for understanding significant relationships among the concepts. It is therefore essential to understand the interactions among the theory of languages, representation theory, and domain theory. Creating a science of EI metrology will require a fundamental and formal approach to measurement methods, testing, and validation analogous to formalisms used in the physical sciences.

Sustainability and Classification of EI Digital Objects

There is a number of potentially useful ways to view the classification of digital objects realized by thousands of digital formats currently in use. One way is through the 3Rs. Another way is by focusing on type of domain. One can also classify digital formats based on whether their content includes a model of the object being represented. For example, a bitmapped image of a part has no object model, but the same part represented using STEP does.

Yet another way is by considering sustainability factors such as those enumerated by the Library of Congress (LC)¹⁹. These are:

- **Disclosure** – the availability of documentation specifying the format and validating software
- **Adoption** – a format’s popularity
- **Transparency** – the ease with which a digital object may be analyzed using generic software (as opposed to specialized tools)
- **Self-documentation** – the inclusion of technical and administrative metadata within the digital object
- **External dependencies** – the degree to which using a digital object requires specialized software or hardware
- **Impact of patents** – the presence of patents related to a digital format
- **Technical protection mechanisms** – the technical methods such as encryption that restrict access to the digital object.

A cursory application of the LC criteria to STEP yields the following observations:

- **Disclosure** – Because STEP is an international standard, documentation is available. However, the official specifications must be purchased from ISO or national standards bodies, and are not available for free on the Web. That said, websites with useful tutorial information exist, but you have to know where to look. Software for validating STEP data also exists, both free/open source and commercial.
- **Adoption** – STEP has been adopted by CAD vendors for exchanging geometry, but has not been adopted as widely for other STEP domains such as product data management.
- **Transparency** – Although it is theoretically possible for a human to analyze STEP data given a physical file and the EXPRESS (International Standards Organization [ISO], [2004](#)) (a language STEP uses for

¹⁹ Sustainability of Digital Formats: Planning for Library of Congress Collections
<http://www.digitalpreservation.gov/formats/>

representing information requirements) schema governing the file, doing so is extremely difficult for a realistically complex data set. Specialized software is needed.

- **Self-documentation** – Because the STEP representation is so rich, digital objects represented using STEP achieve a high degree of self-documentation. Still, it is sometimes useful to add additional annotation using logic-based languages such as OWL²⁰.
- **External dependencies** – Most CAD vendors have a least some ability to import and export STEP geometry data.
- **Impact of patents** – None.
- **Technical protection mechanisms** – None.

To gain an understanding of how best to use the LC criteria to develop sustainability metrics, it would be useful to apply the criteria to other EI formats besides STEP and to look for commonalities and gaps. This would then lead to insights regarding additional EI-specific sustainability factors which, in concert with formal definitions for the 3Rs, would lay the groundwork for the activities described in the following section.

Next Steps

The next steps to developing EI sustainability metrics are to create a framework and test bed based on the DOP work and OAIS, along with the EI modeling requirements derived from our present efforts. The framework should include not only the model for the distributed organization of objects but also the model for creating DOPs in engineering using the information models and an OAIS-based process for defining technologies to create archival packages in the form of DOIs.

The test bed should verify and validate the framework for a product model that covers as much of the product information as possible and is reasonable for testing the framework. The framework will be the basis for designing an EI test bed.

Framework for EI

The EI archival framework should be based on Kahn and Wilensky's DOI methodology as the lowest level model for distributed management, DOP type-object modeling as the next level of detail, and should be combined with product information models and an archival process based on the OAIS standard.

A series of workshops held at NIST and the University of Bath over the last two years provides a source of requirements (Ball & Ding, [2007](#); Lubell, Mani, Subrahmanian & Rachuri, [2008](#); Lubell, Sudarsan, Subrahmanian & Regli, [2006](#)). Workshop discussions attempted to balance the desire for good metadata from those accessing the archive against requirements for ease of use from those populating the archive. The need for both human effort and automation as well as the need for both technology and policies in achieving success should also be addressed. The underlying objective should be to understand the needs of the designated community for EI, and to arrive at a framework for EI using the classification of engineering standards that exist to model information for a product lifecycle.

²⁰ Web Ontology Language (OWL) <http://www.w3.org/2004/OWL/>

The framework design should include a detailed analysis of the LOTAR approach and different example case studies of the OAIS reference model involving other projects in the sciences and engineering. Examples include efforts such as NASA's (National Aeronautic and Space Administration) National Space Science Data Center²¹ for satellite image data and the Centre de Données de la Physique des Plasmas²² for data generated by plasma physics experiments.

The framework should be customizable to different types of engineering information, based on our EI sustainability metrics.

EI Test Bed

The test bed should be a venue for experimenting with and validating the OAIS reference model for archiving EI beyond geometry (to include product structure, assembly, tolerancing, and other areas) based on the proposed methodology and engineering-related archival standards. The suggested approach is to add EI extensions to the OAIS information and functional models and then to implement a pilot archival system employing the extensions.

An OAIS-based information model based on a taxonomy of engineering access scenarios employing the 3Rs should:

- Capture the taxonomy of engineering end-user scenarios
- Define the level and content of a digital object based on the taxonomy of access scenarios
- Classify engineering digital objects based on the 3Rs and sustainability criteria
- Attempt to model the content information in the information model, based on EI sustainability factors

An OAIS-based functional model for EI should address a full range of engineering archival information preservation functions including ingest, archival storage, data management, administration, preservation planning and access. The functional model should:

- Tailor the OAIS functional entities – ingest, archival storage, data management, administration, preservation planning and access – to meet the needs of EI.
- Support preprocessing of the digital information for archiving based on sustainability metrics and driven by the EI-customized OAIS information model. Preprocessing should identify and model all actions undertaken on the submission information package prior to archiving.
- Support post-processing of the digital information, identifying and modeling all actions performed on the archival information package prior to dissemination. Post-processing should be based on content information, metrics, and the end-user requirements.

Using the EI-tailored OAIS information and functional models, a pilot archival information system can then be built employing the 3Rs, sustainability metrics, and the

²¹ National Space Science Data Center, NASA <http://nssdc.gsfc.nasa.gov/>

²² Centre de Données de la Physique des Plasmas Physics Data Centre <http://cdpp.cesr.fr/english/index.html>

DOI and DOP methodologies. The pilot can also serve as a means of determining how easily the digital format identification, validation, and classification tools discussed in the *Related Work* section (fifth strand) can be adapted for use with EI formats.

Expected outcomes of the pilot EI archival system would be:

- A collection of sample data that can be used for testing performance of EI archival system ingest and access
- Identification of designated user committees for EI
- Increased awareness of the need for long-term digital information retention and sustainability

TWR 841 Case Study: A Proposed Experiment

Creating an archive for an existing product with its information scattered and created from a number of sources poses an interesting challenge. The U.S. Navy is currently involved in such an endeavor with the TWR (Torpedo Weapon Retriever) 841.

The Naval Surface Warfare Center Carderock Division (NSWCCD) is creating a collection of digital product models and related information for the TWR 841, a torpedo weapons retriever ship, as part of its Long Term Data Retention of Weapon System Information program (Kassel & David, [2006](#)). This collection is being created from the original drawings dating from the ship's initial design from the 1980s. The collection currently includes product model databases and thousands of files in a variety of formats encompassing 3D product models, CAD data, publications, and drawings. The amount of data and proliferation of formats make it hard to manage and search the collection. NSWCCD in effect is using a collection of systems and tools that allow it to create an integrated environment that spans analysis from early requirements to design of the ship.

The integrated environment includes ASSET (Advanced Surface and Submarine Evaluation Tool) and LEAPS (Leading Edge Architecture for Prototyping Systems) (Hurwitz, [2001](#)). ASSET is a family of interactive applications for use in the exploratory and feasibility design phases of naval surface ships. LEAPS is the product model repository used by the Naval Sea Systems Command. LEAPS enables engineering analysis of a product's design, verifies capabilities, determines manufacturability, and provides value-added feedback. LEAPS supports conceptual and preliminary ship design and analysis integration. LEAPS has programmer interfaces for different kinds of analysis tools, its own internal product metamodel, and indexing mechanisms. The LEAPS metamodel is formally defined and documented using the Unified Modeling Language (UML).

NSWCCD has developed a set of ship-specific use cases to identify the types of queries that their TWR 841 archive should support. Examples of use cases include ship repair, new ship design, cost analysis, bid for construction, and deployment. Examples of information queried include ship compartments, disciplines, systems, zones, and work breakdown structures.

As the TWR 841 project progresses, we have an opportunity to study the various aspects of this integrated system with its variety of legacy formats and systems, large numbers of files, compositions of records, and use cases for future access of the archive.

Our hypothesis is that, by adding packaging metadata to the TWR collection, we can aid the Navy in managing their repository and make it easier to find digital ship information. Further, we believe that by designing a TWR information package schema, we can gain insight into the feasibility of creating engineering-specific customizations of packaging methods developed by the digital library community.

In this experiment we propose to create a packaging metadata XML schema for describing the Navy TWR data and use the schema to create a software application for browsing a subset of the Navy's data collection. The schema can then be evaluated with respect to METS²³, an extensible XML schema standardized by the Digital Library Federation and supported by the Library of Congress for management and exchange of digital objects, and "PREservation Metadata: Implementation Strategies" (PREMIS)²⁴ and/or other schemas approved for extending METS. We focus on METS, both because METS is a widely used standard and also because other packaging methods support METS to some degree, i.e., the METS XML format can be generated from outputs of the other methods. We also propose to use the DOP and DOI concepts in creating archival records using the OAIS process. A demonstration scenario informed by the Navy's use cases can then be developed.

Conclusion

Archiving engineering information and data poses immense challenges in light of technology changes in both hardware and software. Traditional methods that were used in the world of paper are no longer applicable, and a proliferation of digital objects has flooded the engineering workspace in the last 20 years. Initially most engineering information was focused on geometry and some on the development of specific analytical tools, such as finite element analysis and other mathematical models of the artifact being analyzed. During this period, the data formats used have evolved independently to serve specific applications. These developments have led to major problems in maintaining engineering information for products with lifecycles measured in decades. Any technical approach proposed has to be cognizant of the organizational and technological dimensions of the problem. In the engineering-oriented digital archiving workshops conducted at NIST and the University of Bath with wide participation from government, industry, and academia, it was argued that archival models should be such that facilities for archiving are available at the source of information creation. The argument for this approach is that archiving after the fact is often time-consuming and seldom undertaken. This observation and need has important implications for the design of a framework for archiving.

The focus of this paper has been to create a technical approach that takes into account that the problem of long-term EI sustainability starts at the point of creation of the digital objects and ends in delivering the right information for the task at hand (reference, reuse, and rationale) to the end user. The problem is a socio-technical system design problem that requires cognizance not only of the social needs and mechanisms of archiving, but also of the technical possibilities of achieving the archival goals. We have characterized the dimensions of the problem and identified five strands of related work that will inform the design of an archival framework. Last

²³ Metadata Encoding and Transmission Standard (METS) Official Web Site
<http://www.loc.gov/standards/mets/>

²⁴ PREMIS : Preservation Metadata Maintenance Activity (Library of Congress)
<http://www.loc.gov/standards/premis/>

but not least, we have identified a set of issues that need to be addressed in the design and testing of a framework for archiving engineering informatics. We have also proposed how we intend to carry out an experiment with a realistic scenario to explore the issues raised and the nature of the EI framework for archiving. We see this paper as a starting point toward identifying and characterizing the issues pertaining to long-term archiving and management of EI.

Acknowledgements

We are grateful to Ben Kassel, Paul Rakow and their Naval Surface Warfare Center Carderock Division colleagues for sharing their ship product modeling expertise with us and for patiently answering our many questions.

References

- Ball, A., & Ding, L. (2007). *Proceedings of the Atlantic Workshop on Long Term Knowledge Retention*. University of Bath, April 13, 2007. Retrieved November 21, 2008, from <http://www.ukoln.ac.uk/events/ltkr-2007/proceedings/>
- Consultative Committee for Space Data Systems. (2005). CCSDS 650.0-B-1: Reference model for an open archival information system (OAIS). Blue Book. Issue 1. ISO 14721:2003. Retrieved November 21, 2008, from <http://public.ccsds.org/publications/archive/650x0b1.pdf>
- Fenves, S. J. (2001). *A core product model for representing design information*. National Institute of Standards and Technology, NISTIR 6736, Gaithersburg, MD 20899, USA.
- Hurwitz, M., (2001). Information integration via navy LEAPS. *3rd Simulation Based Acquisition Conference*. National Defense Industrial Association, May, 2001. Retrieved November 21, 2008, from <http://www.dtic.mil/ndia/2001sbac/>
- ISO. (2002). ISO 10303-21:2002. Industrial automation systems and integration - Product data representation and exchange - Part 21: Implementation methods: Clear text encoding of the exchange structure.
- ISO. (2004). ISO 10303-11:2004: Industrial automation systems and integration -- Product data representation and exchange -- Part 11: Description methods: The EXPRESS language reference manual.
- Kahn, R., & Wilensky, R. (2006). A framework for distributed digital object services. *International Journal on Digital Libraries, Vol. 6, (2)*, pp. 115-123.
- Kassel, B., & David, P. (2006). Long-term retention of product model data. *2006 SNAME Maritime Technology Conference & Expo and Ship Production Symposium*. Ft. Lauderdale, FL: Society of Naval Architects and Marine Engineers.

-
- Kemmerer, S., (Ed.). (1999). *STEP: The grand experience*. NIST Special Publication 939, National Institute of Standards and Technology, Gaithersburg, MD 20899.
- Kopena, J.B., Shaffer, J., & Regli, W.C. (2006). CAD archives based on OAIS. Proceedings of *ASME Computers and Information in Engineering Conference*, DETC2006-99675, Philadelphia, September 2006.
- Lagoze, C., Payette, S., Shin, E., & Wilper, C. (2005). Fedora: An architecture for complex objects and their relationships. *Journal of Digital Libraries*, Special Issue on Complex Objects, Springer.
- Lubell, J., Mani, M., Subrahmanian, E., & Rachuri, S. (2008). *Long term sustainment workshop report*. National Institute of Standards and Technology, NISTIR 7496.
- Lubell, J., Sudarsan, R., Subrahmanian, E., & Regli, W.C. (2006). *Long term knowledge retention workshop summary*. National Institute of Standards and Technology, NISTIR 7386.
- Saidis, K., & Delis, A. (2007, May). Type-consistent digital objects. *D-Lib Magazine*, vol. 13, no. 5/6. Retrieved November 21, 2008, from <http://www.dlib.org/dlib/may07/saidis/05saidis.html>
- Subrahmanian, E., & Rachuri, S. (2008). Special issue on engineering informatics (Introduction). *Journal of Computing and Information Science in Engineering*, Vol. 8, (1), March 2008.
- Subramanian, E., Sudarsan, R., Fenves, S., Fougou, S., and Sriram, R. D. (2005). Product lifecycle management support: A challenge in supporting product design and manufacturing in a networked economy. *International Journal of Product Lifecycle Management*, Vol. 1, No. 1, 2005, pp. 4-25.
- Sudarsan, R., Baysal, M.M., Roy, U., Fougou, S., Bock, C., Fenves, S. J., et al. (2004). *Information models for product representation: Core and assembly models*. NISTIR 7173, NIST, Gaithersburg, MD 20899, 2004, December.
- Wing, J., & Ockerbloom, J. (2000). Respectful type converters. *IEEE Transactions on Software Engineering*, Vol. 26, (7), pp. 579-593.