Record Management and Design Reuse

Hugh C. Briggs*
Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, 91109

Government mandated records management requirements apply to retention and long term archival of a wide variety of records. Part of the attention is on permanent accession and retention by the U.S. National Archives and Records Administration (NARA) but interim requirements for storage by the source agency are included. As government agencies and the Department of Defense move toward implementations, additional goals often include saving design data for reuse. This paper briefly reviews the government records management requirements then investigates candidate meanings of "reuse" and proposes an enhanced design records retention approach. The recommended strategy that emerges is, for a given program or product family, to invest in rich and readily re-executable preservation of design artifacts for one or two subsequent generations, then downgrade the data in utility through conversions, ultimately reaching the NARA minimum standard for permanent historical-interest archives.

I. Introduction

GOVERNMENT mandated records management requirements apply to retention and long term archival of a wide variety of records. Part of the attention is on permanent accession and retention by the U.S. National Archives and Records Administration (NARA) but interim requirements for storage by the source agency are included.[10] As government agencies and the Department of Defense move toward implementations, additional goals often include saving design data for reuse.

A closely related problem for many design and development organizations is long term retention of design data for operations and maintenance support. Many fielded systems have life spans of 50 years. This is well beyond the availability of the design tools and viewers. This is an emerging motivation for the ISO 10303 Standard for the Exchange of Product Model Data (STEP) family of standards.[13] The STEP standards are an example of operable content, in contrast to computer operating system file format and media standards. Media can have relatively short physical life times and life times of the encoding scheme and file format are determined to a large extent by the economic viability of operating systems. This will be reviewed briefly for background but is not the focus of this paper.

Design information has "value" for reuse in subsequent product development activities that is determined by many factors. Product similarity, as in a product family or program, will be the focus here since the reuse value will most likely be higher than in more clean-sheet innovations. Even so, generational design data from twice prior versions is probably of limited value since the technology, efficiency, cost constraints and functional requirements have probably migrated forward significantly.

Since the need for operational or maintenance support of long lived fielded systems is widely recognized, reuse in new/incremental product design will be the focus here. The requirements for maintenance are similar to new design, in that the engineering activities in repair and modification are similar to those in redesign, but strategies and structures for organizing data retention for design reuse are not as well developed.

Special attention will be given to development of a continuum of data formats and functionality to span reuse value. The resulting requirements for format, media, access and tools will be outlined. This descriptive data is the primary basis for the investment versus return trade at the core of retaining data for design reuse.

The recommended strategy that emerges is, for a given program or product family, to invest in rich and readily re-executable preservation of design artifacts for one or two subsequent generations, then downgrade the data in utility through conversions, ultimately reaching the NARA minimum standard for permanent historical-interest archives.

* Hardware Development Process Engineer, M/S 158-224, and Associate Fellow.

II. Records Management

Materials are the most general grouping of artifacts and may be graphic, pictorial, constructive, or written. Materials are often thought of in terms of entities that have a physical presence, have mass and volume and whose intellectual content might be carried solely in their appearance, size, shape, configuration, arrangement or functionality. Models (physical, not electronic or mathematical), prototypes, and test articles are common examples in design and development. These appear a few times in the government requirements primarily with regards to preservation and retention schedules.

Information is that intellectual content that has meaning apart from the physical object. While information may be expressed in a physical medium, our application here does not differentiate by that medium. Information is a content term usually subclassed by domain or possibly purpose.

Data is information based on observation, measurement, test, or other means. It is often differentiated from more constructed and derived artifacts such as reports, descriptions, meeting minutes, or conclusions. Documents are a specialization of materials which contains written content, independent of media and are sometimes called documentary materials.

Records are materials divided into many kinds, differentiated primarily by ownership, treatment, and retention, rather than content. All material created or received in the conduct of work are records. They are also commonly known as documentation. Non-records are materials such as extra copies of documents kept only for convenience, stocks of publications, books and periodicals intended solely for reference, and personal materials.

Media refers to the physical form and the types of media are usually listed, but the predominant intent is to be inclusive and not allow media to be a differentiator to the meaning of records. Format tends to refer to the internal layout on a particular medium, but, again, the predominant intent is to be inclusive and not allow format to be a differentiator to the meaning of records.

A copy is a duplicate of material, potentially in a different medium. The term is used to distinguish the source from the derivative but is not otherwise a differentiator to the meaning of records.

Information in electronic systems can be records. Final results, reports or conclusions are clearly records. The prior and intermediate states of these eventual final reports are not so clearly called out. An example that is called out as inclusive is records with a change control life cycle that denotes initial and changes. Records are information that can have states such as initial/final, in-process/released, active/inactive, or controlled/uncontrolled. Thus, if it is a record in one state, it is a record in the other states. Non-final data in working files are records.

One of the primary purposes of this discussion is to set retention policies for the various types of records. For example, the NASA Records Retention Schedules are organized by major categories of records such as administrative records, personnel records, property records and transportation records.[8] Of course, program and project records are a category and are listed in Chapter 8.[9] This contains many pages of retention schedules for the line items 8000 - 8999, but the best information about the meaning of records in this class is contained in the Items 101 to 113 and the Notes 1, 2, and 3. This chapter starts with definitions of drawings and includes a definition of model that represents architectural scale models rather than computer-based 3D computer aided design (CAD) models. In hardware engineering, similar scale models are common too, but “models” are predominantly CAD solid models. The life cycle stages are formulation; approval; design development; manufacture, fabrication and assembly; pre-launch system integration and verification; implementation and operations; observation; evaluation and termination.

Retention indicates the required period of time the records must be maintained by agencies and contractors. Permanent retention indicates that the records are transferred to a NARA sanctioned records center for storage after project close. Certain temporary records may be destroyed when no longer needed and others are designated to be retained between 5 and 30 years after project termination. The length of this period depends upon the data’s operational value and the project’s need for operational support.

More recent activities are also based on these regulations. For certain government contractors, material subject to disclosure under the Freedom of Information Act may be managed separately from other material. Much of the material generated in process of doing work is thought to be of value to future work. This notion of “reuse” is the primary topic of this paper.

The majority of these records are for historical or auditing purposes and aren’t of much help in reuse. The regulations are clear that the purpose is long term historical preservation along with an emerging need to provide visibility and accountability of operations. The following sections illustrate the limitations of such simple archival activities and sketch a classification scheme for information of more use in subsequent project design activities.

III. Knowledge Management

Enterprise Knowledge Management (KM) projects are a common way for many organizations to deal with pressures from regulations such as Sarbanes-Oxley, from loss of employee experience through retirement of the Baby Boomers and from quality improvement and certification processes.

Knowledge Management Systems (KMS) have been defined as a line of systems which target professional and managerial activities by focusing on creating, gathering, organizing and disseminating an organization's "knowledge" as opposed to "information" or "data."[1] Most organizations pursue KM through a blend of approaches, with designated roles and departments implementing KM.[16] Knowledge Management literature casts a firm's knowledge assets as its structure, culture, processes, employees and physical artifacts, with an emphasis on employee knowledge.

KM has progressed to become an enterprise activity known at the executive level.[16] It is postured as a business critical information activity. Its business case is based on effective practices for "knowledge workers," aimed at minimizing information recovery and rediscovery and effective collaboration for team members.[5] A second motivation is to enable the enterprise to more effectively handle the transition of the work force. Two transitions at the forefront of the argument are 1) the retirement of the Baby Boomers and 2) the increasingly migratory situation for noncraft employment. To be an effective Information Technology (IT) system for a learning organization, the organization might also need to reengineer its business processes.[4]

Typical implementations capture work activity communications such as email,[14,15] electronic memoranda and business documents such as orders, proposals, position papers, etc. Content Management (CM) is defined as software that builds, organizes, manages, and stores collections of digital works in any medium or format.[6] Many implementations try to mimic and seek to replace existing storage services for these expressions, such as the PC hard drive, the department file server and the email server. To maximize adoption and minimize manual front side effort, these implementations want to be drop in replacements. More complex systems integrate with the authoring environment to extract metadata from documents and classify them as they are processed.[6]

The broader notion of "knowledge management" is greater than simply taking care of knowledge that exists, or transforming from one form, such as the implicit or tacit knowledge of experienced personnel, to another form, such as the explicit knowledge in guidelines and manuals.

Knowledge Management can be broken into four stages that require different activities.[2]

Knowledge Capture is the identification of structure and the collection of data into that structure. This might occur in special activities aimed at developing structure or organizing data, or it might happen implicitly during the accomplishment of an activity. The result is a data store with an organization that is of greater value than the isolated facts.

Knowledge Management, as a subtask, is the maintenance, indexing, cleaning, preserving and persisting of the knowledge store. In itself, the activity doesn't add more knowledge, but the maintenance processes might make the existing knowledge more valuable or more usable. The construction of indices to assist in searches is an example of adding value.

Knowledge Access is the set of methods for inquiring of or investigating the knowledge store. The knowledge store organization can be complex and finding information, drawing conclusions or identifying patterns requires mechanisms for finding, collecting, aggregating and displaying desired information.

Knowledge Application is the end act of using the knowledge found in the store in a current activity. This is the purpose of Knowledge Management and the reason for collecting the information, maintaining the store, and studying its contents.

Within this structure we can derive requirements for saving project engineering records and data. For example, integrated search engines provide access. Reusing discovered data to assist in new designs, the core of "reuse", is knowledge application. The remainder of this paper investigates further the impact of discovery and application on the type and format of saved records.

IV. Integrated Search

Significant development effort has been made lately in integrated search across disparate data stores. An essential element of this is the information structure consisting of metadata and its list of values. Successful search and discovery is dependent on quality and organizational structure of the descriptive metadata.

The Resource Description Framework (RDF) is a W3C Recommendation for describing resources for use in a variety of processes and is meant to be used in activities such as resource discovery, intelligent software agents, cataloging, and the semantic web.[17] The RDF data model entails three main elements: resources, properties and relationships. A domain ontology is an abstracted collection of entities with properties and relationships. A taxonomy is the controlled vocabulary listing the allowed entities and values for their properties. This data structure provides the organizational scheme for unified or integrated search capability across disparate collections.

Integrated search provides the means for project users to find and retrieve archived data from prior projects.

Project data are most likely stored in multiple IT systems with differing specializations and search and investigation schemes. For example, released drawings for prior project flight hardware will be in the organization's enterprise release management system. This system will store images of the drawings along with metadata about its approval, change process, and adoption for use on projects.

Design review material will be vaulted in a project documentation library which may be more like a document management system. The system will store the original document as a Microsoft PowerPoint or Word file, metadata about the author and data about its subsequent changes, if any.

Integrated search will assist the next project design team in finding a wide variety of data from all of the organization's IT systems. For example, a search for information on a gear box assembly should find the released drawings, the design review material and other analyses.

V. File Formats and Media

Archivists have come to a good understanding of media lifetimes, if one distinguishes every day or casual back ups from archival managed copies. In the latter case, care and a favorable environment can easily yield 10 - 30 year lives for recorded magnetic media and 70 or more year lives for recordable compact disk and digital video disk media.[18, 19, 20, 21, 22, 23] Best practices for retention include scheduled retensioning, recording refresh, and transcription. System technology is ultimately the limiting factor and technology migrations are part of a sound retention strategy.[21]

File formats are another issue to be dealt with, beyond the issues of character set and end of line marks. The basic text file is entirely too poor to be of any help. Richer content file formats become tied to viewer programs.

Most project documentation in the US aerospace and defense companies is in Microsoft Office. Presentations are in PowerPoint and memos and reports are in Word. Project documentation is often archived for future reference in non-editable forms such as PDF.

Archivists have struggled with the question of allowable formats, or equivalently supported viewers. For example, NARA accepts images in a few basic formats: TIFF, GIF, and JPEG.[11] Much of the NARA direction is aimed at the minimum resolutions and lossless encodings required to facilitate Optical Character Recognition (OCR) for text recovery.

NARA has recently accepted PDF format documents.[12] All PDF versions at the time (v1.0 through v1.4) were accepted and the discussion concentrated on advanced topics such as security (no embedded passwords), fonts (all fonts required must be embedded) and scanned documents (again, resolution sufficient to support OCR text recovery).

Beyond such basic archival format and viewer examples, the most prevalent format for mechanical engineering design files is the STEP Standard (ISO 10303). Recent initiatives such as the Long Term Knowledge Retention[24, 25] and the Long-Term Archiving and Retrieval[26] projects are extending the basic STEP schema to represent the explicit model accurately in a product neutral format. This will improve the quality of the model by capturing the design intent or design processes in support of regulatory and contractual compliance and design re-use.[13, 26]

VI. Product Life Cycles and Information Value

A. Design Data Utility

Saved information can be useful in various ways. This section surveys briefly the ways that information can be used and then ties these to the form of the information. Much of the electronic data available today is, of course, potentially available in various forms. The premise of this paper is that those organizational leaders who are designing record retention practices that facilitate later reuse can choose the form that best fits the organization's investment potential.

System design is both a thoughtful and an active endeavor. It is done by individuals, collocated groups and geographically separate groups. Designers use information across a continuum of interaction: awareness, survey, replay, recomputation, and extension.

The most basic need might be awareness of prior work. With awareness, the wasted resources of time and effort in the rediscovery and reaccomplishment might be avoided. It might be enough to view a summary of the resulting system design in the form of figures, drawings and illustrations, but these are more valuable when placed in the context of the system objectives and requirements. That is, a summary of both the problem posed and the answering design is substantially more helpful to subsequent design inquiries. This information might be preserved in a project summary report, a system design review or journal articles. Additional depth for subsystem designs saved in similar forms would provide help for the technical disciplines.

Beyond a statement of the question and its answer, the documentation might discuss the path of discovery taken, the choice criteria that guided selections and the deficiencies of the alternative designs not chosen. Such information is often included in the design reviews that conclude these activities. Future work in similar system designs might pursue a non-chosen design, or might make the choices differently due to new circumstances.

Current record management practices are focused on these documents, which might be called project documentation, when design reuse is added to the NARA requirement for historical preservation. For example, JPL hardware development procedures define controlled records and enumerate the design memos, reports and review materials, across the project life cycle that must be preserved.[7,3]

Clearly, there is benefit to future project design work in capturing such documentation provided it can be found and studied effectively. This is the goal of the JPL unified search activities and the Records Management Initiative. With this in place, future project teams can be aware of relevant prior design work and to a limited extent understand how the design works and why the prior team made it that way.

Engineering design activities use a wide variety of tools beyond the final report and presentation. Visualization in line drawings, CAD models, animations, and sequences are significant contributions because of the importance of vision and spatial layout in design. Numerical models, such as dynamics simulations, thermal analyses, flight performance simulations, flutter prediction, mission simulations, fuel economy and range analyses, and stress analyses, are used to predict technical performance. The product configuration, layout, articulation motions, and interferences are analyzed with CAD models. The performance of the control system, data management system and electronics is analyzed with circuit, network and software simulations. The results and conclusions of using these tools are documented in reports that are frequently preserved in project documentation.

It is at this point that engineering management falters in defining preservation practices beyond documentation. It is clear that preserving this type design data in reusable, specifically re-executable, form requires preservation of both the analysis tools, the computational environment and the computer hardware. Only in the most extreme situations has this been practical. The executable lifetime of hardware and operating systems is perhaps five to ten years. Most commercial analysis software supports only one or two prior versions, which can result in model lifetimes of 3 to 4 years.

This quandary is widely understood and is a significant motivation for many standards. STEP is a mature neutral standard format for many engineering design models including CAD, analysis and requirements. As is the nature of standards, quite a bit of detail design information is lost when exporting proprietary CAD models to STEP, but the geometry description and layout are accurately captured. As mentioned, the STEP standards are being extended to retain more of this information.

For structural analyses, the NASTRAN data deck is a de facto standard with significant life. Like a STEP CAD model, it provides an accurate performance simulation of the design, but is not well suited for design exploration. The data deck is an intermediate product in modern tool chains and analysts work principally in preprocessors while abstracting design models and preparing the data deck and in post processors while investigating, formatting and presenting results.

Thus, while STEP models or analysis decks capture a specific design, neither provide a good basis for new design development.

B. Program Life Cycles

New product designs frequently respond to new requirements, or take advantage of new technology and, as a result, are extensions, evolutions or extrapolations of current designs. It’s common to manage this series of products as a program of sequential related projects. It is in this context of sequential design that this paper seeks to help motivate the investment decision.

In programs, then, the most relevant design data is that of the current design. The design prior to that is sometimes helpful, if only to put the current design into its proper context. The rate of change of engineering tools, cultural structures, market pressures, and component technology clearly impacts, for each program line, the relative worth of prior product design information.

Each program will have its own generational pace, component technologies and business value. The generation-to-generation time compared to viable life times of executable saved design data provides a technical limit to the value of design “reuse.” Clearly, a rapidly evolving family of products, say with one or two year product updates, won’t be restricted in this way.

A program office whose products have substantial future sales prospects and margins will have the financial means to support future design teams with rich extensive executable archives. On the other end of this spectrum, a competitive, cost constrained market place will limit resources for retention.

This equation is substantially more complex in government programs with certain financial accountability requirements. For example, funds to build a given project cannot be used for another project. Consequently, program managers and organizational mangers will need to develop funding sources that have the future projects at heart.

VII. Recommendation

This, then, provides the basic structure for both sides of the investment question. The choice of preservation form sets the costs and the value to subsequent product design sets the return.

For each program, engineering management and program management must determine the generational time lines and value of design elements by subsystem, technology, and discipline.

For high value product lines, the strategy could easily be:

- Retain design models in active supported commercial tools and compute environments for use in the next generation product design.
- For the subsequent generation, port the data to standard formats executable in compatible commercial tools.
- For longer term retention, preserve the data in viewable but non-executable forms such as project documentation.

For lower value, or more slowly evolving programs, the strategy might be:

- Capture the final design in standard formats and expect limited reuse as commercial compatible tools progress onward.
- Capture the project documentation at project conclusion with good thoroughness and expect this to be the primary basis of reuse.

This strategy requires a few new practices as well as an investment during the current project that is to the benefit of the future project. For example, the strategy for high value information requires the down conversion from commercial tool formats to standards-based formats a generation and a half after the project completion. As a result of accounting rules, management will likely need to invest internal funds in the production and preservation of such data during the final stages of project completion.

VIII. Conclusion

Current archival record management notions of retaining project documentation are clearly insufficient to provide substantive help to future projects. Given success of the new integrated search tools, future teams will be better able to find project documentation, but the design value in the viewables will be quite low. The business case for preserving design data in more valuable formats is easy to formulate, but certain required values aren't well known. For example, how much would a future project team save if they had better data? How much would it cost to collect, preserve and transform such data across two or three project generations?

As result, this strategy is best tried first on high value projects where risky investments can be protected.

Acknowledgments

This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

References

1. Benbya, Hind and Belbaly, Nassim Aissa, "Mechanisms for Knowledge Management Systems Effectiveness: An Exploratory Analysis," *Journal of Knowledge and Process Management*, Vol. 12 No. 3, 2005, pp 203-216.

2. Briggs, H.C., "Knowledge Management In The Engineering Design Environment," Paper AIAA-2006-2238, 47th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics & Materials Conference, Newport, Rhode Island, 1 - 4 May 2006.

3. Briggs, H.C., “Managing Analysis Models in the Design Process,” 48th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics & Materials Conference, Honolulu, HI, 23 - 26 Apr, 2007.

4. Chen, Jim Q., Lee, Ted E., Zhang, Ruidong, and Zhang, Yue Jeff, “System Requirements for Organizational Learning,” Communications of the ACM, Vol 46, No 12, Dec 2003, pp 73-78.

5. Davenport, Tom, “Knowledge Management Round Two,” CIO Magazine, 1 Nov 1999.

6. Krishna, Vikas, Deshpande, Prasad M., and Srinivasan, Savitha, “Towards Smarter Documents,” ACM CIKM’04, November 8-13, 2004, pp. 634 - 641.

7. Linick, T.D. & Briggs, H.C., “Developing the JPL Engineering Processes,” Paper AIAA-2004-6129, 45th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics & Materials Conference, Palm Springs, CA 19-22 Apr, 2004.

8. National Aeronautics and Space Administration, NASA Records Retention Schedules, NASA Procedural Requirements, NPR 1441.1D, http://nodis3.gsfc.nasa.gov/displayDir.cfm?Internal_ID=N_PR_1441_001D_

9. National Aeronautics and Space Administration, NASA Records Retention Schedules, Chapter 8, NPR 1441.1D, http://nodis3.gsfc.nasa.gov/displayDir.cfm?Internal_ID=N_PR_1441_001D_&page_name=Chapter8

10. The National Archives, http://www.archives.gov

11. National Archives and Records Administration, Frequently Asked Questions About Imaged Records, http://www.archives.gov/records-mgmt/faqs/imaged.html

12. National Archives and Records Administration, Expanding Acceptable Transfer Requirements: Transfer Instructions for Permanent Electronic Records, Records in Portable Document Format (PDF), http://www.archives.gov/records-mgmt/initiatives/pdf-records.html

13. Pratt, M., “STEP and Related Approaches to Long Term Knowledge Retention,” Atlantic Workshop on Long Term Knowledge Retention, Bath, UK, 12 – 13 February 2007.

14. Smallwood, Robert, “E-mail management hits center stage—Part 1,” KMWorld, Volume 14, Number 3 March 2005.

15. Smallwood, Robert, “E-mail management hits center stage—Part 2,” KMWorld, Volume 14, Number 4, April 2005.

16. Vestal, Wesley, ”Making sense out of KM costs,” KMWorld, Volume 14, Number 7, July/August 2005.

17. World Wide Web Consortium (W3C), Resource Description Framework (RDF), http://www.w3.org/RDF.

18. The CD-Info Company, “TDK CD-R technology,” http://www.cd-info.com/CDIC/Technology/CD-R/Media/TDK.html.

19. Stinson, D., Ameli, F. and Zain, N., “Lifetime of KODAK Writable CD and Photo CD Media,” Eastman Kodak Company, 1995, http://www.cd-info.com/CDIC/Technology/CD-R/Media/Kodak.html.

20. Gilbert, M.W., “Digital Media Life Expectancy and Care,” Office of Information Technologies, University of Massachusetts Amherst, 2003, http://www.softpres.org/cache/DigitalMediaLifeExpectancyAndCare.html.

21. Van Bogart, J.W.C., “Magnetic Tape Storage and Handling, A Guide for Libraries and Archives,” National Media Laboratory, June 1995, http://www.clir.org/pubs/reports/pub54/index.html.

22. Puglia, S., “Preservation Reformatting: Digital Technology vs. Analog Technology,” 18th Annual Preservation Conference, National Archives at College Park, Maryland, March 27, 2003, http://www.archives.gov/preservation/conferences/papers-2003/puglia.html.

23. The National Archives, “Frequently Asked Questions (FAQs) about Optical Storage Media: Storing Temporary Records on CDs and DVDs,“ http://www.archives.gov/records-mgmt/initiatives/temp-opmedia-faq.html.

24. ProSTEP, “Long Term Archiving and Retrieval of Product Data within the Aerospace Industry (LOTAR),” http://www.prostep.org/en/projektgruppen/lotar.

25. ProSTEP, “White Paper for Aerospace Long Term Archiving and Retrieval,” August 2002, http://www.prostep.org/file/18934.WP_LOTAR.

26. PDES, Inc., “Long Term Data Retention (LTDR) Pilot,” January 1005, http://pdesinc.aticorp.org/long_term_data_retention_pilot.html.