Computer-Aided Design 43 (2011) 820–837

Contents lists available at ScienceDirect

Computer-Aided Design

journal homepage: www.elsevier.com/locate/cad

On the long-term retention of geometry-centric digital engineering artifacts

William C. Regli *, Joseph B. Kopena, Michael Grauer

Department of Computer Science, College of Engineering, Drexel University, Philadelphia, PA, USA

ARTICLE INFO

Keywords:
Digital preservation
Long-term knowledge retention
Representation
Standards
Knowledge capture

ABSTRACT

This paper discusses the challenges of long-term preservation of digital geometric models and the engineering processes associated with them. For engineering, design, manufacturing, and physics-based simulation data this requires formats that are accessible potentially indefinitely into the future. One of the fundamental challenges is the development of digital geometry-centric engineering representations that are self describing and assured to be interpretable over the long lifespans required by archival applications. Additionally, future users may have needs that require other information, going beyond geometry, be also accessible to fully interpret the model. These problems are highly interdisciplinary and not exclusively algorithmic or technical. To provide context, the paper introduces a case study illustrating an overall portrait of the problem. Based on observations from this case study, we present a framework for enhancing the preservation of geometry-centric engineering knowledge. This framework is currently being used on a number of projects in engineering education.

1. Introduction

Digital Preservation is the mitigation of the deleterious effects of technology obsolescence, media degradation, and fading human memory [1]. One of the fundamental challenges facing those developing Digital Preservation solutions is the development of digital representations that are self describing and assured to be interpretable over the long lifespans required by archival applications.

This paper describes challenges related to the long-term retention of digitally represented engineering artifacts. In this context, a *digital engineering artifact* refers to an object that was designed, modeled, analyzed, simulated, manufactured or maintained principally "in silico" using software tools (e.g., CAD, CAM etc.) rather than via paper or physical prototypes. Nearly all contemporary engineering artifacts, whether automotive, aerospace, architectural or other consumer goods are at least, to some degree, digital artifacts. Digital Engineering Archives are the digital repositories for engineering design, manufacturing and life-cycle data.

As engineering enterprises complete the transition, begun in earnest at the end of World War II, from trades-craft to information-centric disciplines, the fundamental challenge has emerged as to how to ensure the archival and transmission to future generations of the engineering knowledge that is immensely complex, inter-dependent, interdisciplinary and now encoded in a myriad of digital file formats and software systems. Digital Engineering Archives require the development of file formats that can capture the necessary and sufficient knowledge to ensure future interpretability and be physically accessible "in the future". What "in the future" means may vary by industry, for example the aerospace and ship building industry work around 30-to-50 year lifespans; architecture-engineering-construction (AEC) deals potentially in centuries. What makes engineering data particularly complex to preserve is the complexity of individual file formats formats and vast set of metadata required to understand how all of the aspects of artifacts are connected.

To better frame the set of issues surrounding these problems, this paper contends that the challenge of digital engineering archives is fundamentally one of *representation*, specifically representation of engineering information and knowledge at four distinct (but highly inter-related) levels:

- *File formats*: All digital objects to be preserved are required to have a representation that can be written to persistent storage, such as a disk, tape or other medium. All engineering software systems have the means to save files in this manner, however many of these formats are proprietary, often intentionally designed to limit interoperability. Further, the diverse set of tools and activities in any engineering organization can result in a constantly changing set of relevant formats, files and documents relevant to a given artifact.

  A challenge is how to track these formats, understand their information content and role in interpreting an artifact.
- *Logical object encodings*: Each of these file elements has an encoding describing the semantics of their contents. This certainly includes, but is not limited to, the grammar with which a file is parsed and interpreted into an internal data structure by a piece of software. While in some ways the logical

* Corresponding author. Tel.: +1 215 895 6827; fax: +1 215 895 0545.
E-mail address: regli@drexel.edu (W.C. Regli).

0010-4485/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.cad.2010.11.012

object encoding can be considered the syntax associated with a file format, the semantics of such an encoding is not necessarily explicitly represented and left open to interpretation. These challenges are particularly acute for the geometry-centric representations at the core of computer-aided design where, in spite of rigorous underlying mathematical formulations [2], the interpretation of geometry/topology relationships [3–7], as well as the ambiguity of feature-based [8–11] and parametric models [12,13], has been widely studied over the past three decades.

This challenge is one that has been taken up most strongly by the standards community and in the engineering context by the ISO Standard 10303, known as STEP [14].

- *Object metadata*: What is becoming increasingly clear in digital organizations is that having self-contained files alone is not enough. Issues of version control, object provenance, traceability and document history are critical for a wide variety of purposes, including forensics, accountability, diagnostics, and life-cycle maintenance. While representations such as STEP have begun to include feature and construction history as part of the internal logical semantics, this does not provide enough to ensure the "system"-wide needs for data-driven engineering are met. For example, there may be multiple nearly-similar copies of a digital object within an organization—which one is the "current" or master version? How do the derivative versions relate to each other? Which are essential and which are redundant; and for what purposes were the derivative versions created? In some cases, surrogate models are needed for performing analysis, simulation or to provide to subcontractors.

  Maintaining associations across all of these data models poses a well-known set of open problems that are currently intertwined with collaborative work systems and Product "Life-Cycle" Management products (PLM) [15–17]. However, typically today, these metadata objects are not external to the systems the systems that track them, nor are there established standards for many of the elements.
- *Organizational workflows*: A workflow describes the sequences of actions or operations, taken by an individual or a group, involving an engineering artifact. The study of workflows has been around for decades and has been studied in the context of engineering organizations in a wide variety of ways. Typically, workflows are used to describe business processes or to develop models for understanding organizational behavior.

  While there are many representation techniques for workflows, ranging from the IDEF family of modeling languages to scientific workflow systems such as Taverna and Kepler, few of these are designed for preservation purposes or tightly coupled to the engineering domain.

Noting the above four categories, the majority of recent work in digital preservation for engineering has centered on logical encodings of individual files and the enhancement of the STEP standard to serve this purpose [18–21,13,22–24]. This stems in part from the belief that archiving engineering objects can proceed in a similar manner to that of other media (i.e., film, audio, images). Archiving digital engineering data presents additional problems beyond those of archiving in general, such as the mathematical complexity and the proprietary nature of many CAD file formats, as well as the need to capture the organizational workflows to understand the rationale behind the models one may have. Unlike "traditional" digital media, engineering data is largely useless without the broader product model context that describes the 'how' and 'why' and 'who' behind it. For example, the primary goal of an archive of an audio track is to support those that wish to analyze, study or simply listen to the audio. In contrast, the goal of the engineering archive is to support future users and enable them to understand the model. In this case, "understand" implies concepts that go beyond just a rendering of the shape or a 2D blueprint and includes the model's intended use, its design process and manufacturing constraints, and possibly much more. Since so many of the processes and data formats involved in engineering design are complex and opaque, it becomes necessary to develop knowledge structures that can be used to archive a wide variety of individual data formats and capture the vital inter-relationships needed to understand the engineering records.

The needs of digital preservation for geometry-centric objects are somewhat different than those that have driven traditional research advancements in geometric representation, advancements that usually are directed at problems of mathematical accuracy or application end goals (i.e., improving machining, performing FEA, etc.). The fundamental property of a preservable geometry object is to enable information exchange across temporally distinct software epochs. In this way preservation is similar to work on data exchange, which focuses on translation between different software systems or across incremental version releases of the same software (i.e., data migration). Representations suitable for digital preservation need to inter-operate with "future" software systems (i.e., those yet to be implemented) and cannot assume that these systems share much, if any, similarity with existing tools. Depending on the domain, the temporal epoch could be measured in decades (i.e., for aerospace artifacts) or centuries (i.e., for architectural and civil structures)—thus ensuring that the very existence of corporations and software vendors cannot be assumed.

This paper has several goals and audiences:

- Beyond just engineers, there are archivists, librarians, artists, information scientists, historians and many of other disciplines are becoming increasingly involved in the curation and stewardship of digital engineering data. This paper offers a high-level overview of many of the challenges and issues that might not be familiar to those not grounded in the history of CAD and 3D model formats.
- We present a framework in which geometry-centric artifacts and their associated physics-based models, manufacturing models, simulations, etc. can be captured and packaged with a reasonable guarantee of future interpretability. Our approach is to take representations for geometry-centric preservation and augment them to include semantics beyond the geometry itself. For example, some geometry-centric representations are procedural or feature-based, resulting in the need to also understand how to interpret the execution of the process for generating the shape. In these cases one would want to preserve both the end shape result as well as the generative process.

  Future interpretability of geometry-centric objects also requires some capture and association of *context*. For purposes of this work, *context* describes the use, functions or process associated with a geometry-centric artifact that are essential to maintaining its future utility. For example, for a future user to understand an individual discrete part might also require some semantic model of its role in a larger assembly; its manufacture may require having saved models that describe intermediate "in-process" shapes as well as jigs and fixtures, etc.
- To ground this work, the paper provides real-world case-studies in the form of the AMBER part and the Cyber-Infrastructure-Based Engineering Repositories for Undergraduates (CIBER-U) Collaboratory [25]. The Advanced Model-Based Engineering Realization (AMBER) part is a test artifact from an automated manufacturing demonstration by the Ministry of Defence of the United Kingdom. Describing just this single part required over 3.5 GB of data, including 39 file formats and over 750 distinct model and shape files. Our belief is that the case-study represented with AMBER is commonplace in engineering practice today and lessons learned from this dataset may be