Motivations and Challenges for Digital Preservation in Design and Engineering

Dominic Heutelbeck, Jörg Brunsmann, Wolfgang Wilkes
Faculty of Mathematics and Computer Science
University of Hagen
D-58084 Hagen
+49-2331-987-2172

Andreas Hundsdörfer
InConTec GmbH
Kirschenallee 7
D-96152 Burghaslach
+49-9552-931494

dominic.heutelbeck@fernuni-hagen.de
joerg.brunsmann@fernuni-hagen.de
wolfgang.wilkes@fernuni-hagen.de
andreas.hundsdoerfer@incontec.de

ABSTRACT

Digital preservation is a fairly new consideration in design and engineering companies. The core motivations for enterprises to engage in digital preservation endeavors are of economic and legal nature. This paper provides an overview of software components and processes in typical engineering scenarios, based on electrical engineering as an example. In addition, it describes typical product data in an engineering process which is subject to potential archiving. Based upon the engineering scenario and motivations, the challenges for digital preservation are derived and examined. These challenges cover a wide range of software engineering and scientific challenges.

Categories and Subject Descriptors

J.6 [**COMPUTER-AIDED ENGINEERING**]: Computer-aided design (CAD) – *digital preservation***.**

General Terms

Documentation, Design, Economics, Reliability, Human Factors, Standardization, Legal Aspects

Keywords

computer-aided design, digital preservations, metadata, archives, engineering processes, reuse

1. INTRODUCTION

The notion of digital preservation in the form implemented by memory institutions is fairly new to companies and institutes working in the engineering domain. For classical participants in digital preservation activities, such as memory institutions and museums, the need for digital preservation with the perspective of archiving data for more than a hundred years is obvious. In fact, long term preservation is their primary objective and digital preservation is a natural extension. This objective is also the reason for many of these institutions to exist at all. However, the motivation for engineering companies is different. The primary objective of most of these companies is to be a profitable business within their domain. It is important to examine how digital preservation can support businesses. Section 2 describes which motivations drive engineering companies to participate in digital preservation activities. To approach the domain of design and engineering we take a closer look at electronic product design and the associated product lifecycle in section 3. When designing digital preservation systems, thorough knowledge about the artifacts which are to be preserved is critical. The documents and data associated with a product design and lifecycle are called **product data**. Section 4 describes product data by highlighting key properties. Based upon the observations made in the previous sections, section 5 presents challenges for digital preservation systems supporting design and engineering. Section 6 concludes with a short summary and the authors' view on how these challenges can be most fruitfully addressed by industry and researchers in the future to build a new exciting domain for digital preservation.

InDP'09, June 19, 2009, Austin, TX, USA.

2. MOTIVATION

For establishing digital preservation in design and engineering companies, there should be clear economic reasons for a company to engage in preservation activities.

In addition to possible economic benefits, in certain scenarios, there are clear legislative motivations for investing in digital preservations. For example, law and individual contracts may require certain types of technical documents to be archived for at least 50 years.

In the following, we call all data which are potentially subject to archival product data. This includes CAD designs, technical documentation, metadata and much more, as described in sections 3 and 4.

In this section we examine, which economic and legal reasons are relevant for individual businesses within the engineering domain.

2.1 Legal reasons for digital preservation

There are at least three different types of legal reasons for archiving product data, e.g., designs, documentation, audit trails, or simulation results:

- Archiving ***required by law***. In some scenarios, such as medical equipment for the US market, the FDA Part 11 regulations [19] have to be fulfilled.
- ***Contractual requirements***. When designing products with a long lifetime, the client may require that detailed technical documentation shall be available, e.g., for maintenance. For example, systems designed for the military sector often serve more than 50 years, and the contracts are designed to accommodate this fact. Contractual partners require that other partners implement certain quality standards, such as ISO 9000ff [8].
- Preparing ***legal defense***. A complete audit trail of documents can possibly strengthen the position and provide evidence in a lawsuit.

If these requirements, i.e. by law and/or by contract, apply, archival is mandatory. If a company operates within scenarios, where these requirements apply, advanced preservation technologies can both enable the fulfillment of the requirements, as well as potentially reduce the cost for the fulfillment compared to systems not designed to be integrated with preservation systems.

We suggest that companies carefully review the legal requirements within their projects and consider preservation technologies accordingly.

2.2 Economic reasons for digital preservation

Beyond the legal requirements, we believe that deploying digital preservation technologies has potential for actual economic advantages. For example, as stated in section 2.1, implementing proper preservation strategies may reduce the cost associated with the mandatory legal requirements.

Furthermore, engineering companies are steadily increasing their assets through innovation. Securing and preserving these assets is in essence what digital preservation is all about. Thus, digital preservation can help in securing assets for long time exploitation. Assets may include all product data as well as more general knowledge, intellectual property, and patents. The following sections highlight different aspects and how preservation aids in exploiting assets.

2.2.1 Reuse

Competition drives business to lower the development costs while reducing time-to-market of products. In order to save development costs and to deliver new products faster, a main objective of companies is to reuse existing designs. One of the core premises of current research in digital preservation is, that while data may be archived and reproduced easily by copying bits, ensuring that the data is usable in the long run is a hard problem. The reason for this is that technology is advancing quickly. The lifespan of a design or product may exceed the lifespan of several generations of engineering tools, operating systems and other related software products, i.e., the engineering environment. Thus, it is not guaranteed, that the product data, i.e., CAD objects, are useable and reusable in the latest engineering environment. This is exactly the kind of problem, current research in digital preservation, e.g., the SHAMAN project [17], try to solve by providing different technologies for managing the evolution of systems while maintaining the integrity, semantics and usability of data. Thus, using such technologies, **enable to reuse designs** beyond the lifespan of a single engineering environment.

In addition, we assume that archiving designs actually **increases the probability of reuse**. The reasoning behind this assumption is that by implementing digital preservation strategies, the effort for reusing product data is reduced to a point where reuse becomes economically viable. As a consequence, one can think of the **assets becoming more valuable** through the application of preservation technologies. Since we assume that the probability of reuse decreases with time, it is necessary to have high availability (documentation, finding) in the present.

A specialty of CAD objects regarding reuse is that reuse can be performed on different levels:

- **Complete reuse**: Complete design reuse in a new system.
- **Partial reuse**: Reuse of a part of a design in a new system.
- **Functional reuse**: Instead of reusing a design completely, only the description of the functionality of the design is reused to reimplement the functionality with new technologies.

2.2.2 Audit and Investigation

In certain scenarios, it is important to be able to reconstruct decision making and design processes. Motivations for this are:

- **educational**, e.g., someone needs to learn how to execute a certain task that has been performed before
- **investigational**, e.g., someone needs to find out why a decision was made or who is responsible for it.

Digital preservation can support this by improving traceability through archiving the matching audit trails.

2.2.3 Maintenance and repair

Maintenance is both relevant for the design itself for correcting errors, as well as for actual physical implementations of designs, i.e., machines built to the specs of the design, when parts become defective. To be able to perform maintenance, the matching documents have to be available during the entire lifespan of the design or its physical implementations. Maintenance takes place on different levels

- **Chip level**: individual components on a board.
- **Board level**: individual printed circuit board in a system.
- **System level**: system including cabling and enclosures.

2.2.4 New requirements

According to the product planning of a company, existing product data, such as CAD objects and their implementations, need to be extended or rebuild according to new requirements. This includes new innovations or variations of existing products. For these tasks, the designs have to be available in a similar manner as for maintenance. Especially it has to be ensured, that it is possible to change the design.

requirements for digital preservation environments. While doing so, it is necessary to consider the whole product life cycle and identifying all involved participants of that product lifecycle including their roles and responsibilities.

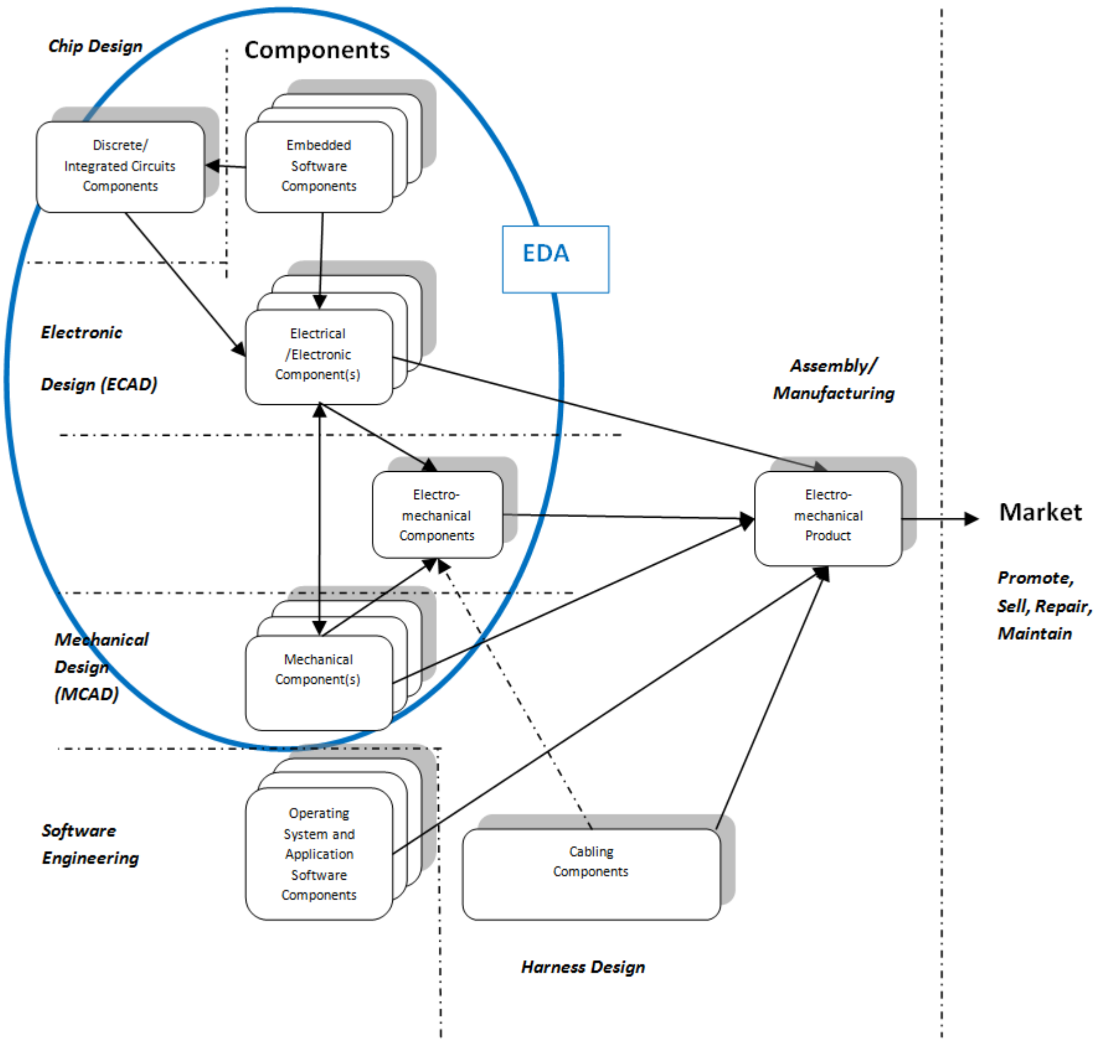

Figure 1: High-level view of product composition

2.3 Overview

We can conclude that it is possible to categorize the motivation for digital preservation into legal and economic motivations.

In addition, one should note, that motivations relate to business activities which are oriented differently regarding their time perspective:

- Past (Audit, Investigation)
- Present (Availability, Productivity)
- Future (Reuse)

As described above there are numerous motivations for engineering companies to invest in digital preservation. The given motivations can be used to derive specific use-cases and

3. ELECTRONIC PRODUCT DESIGN AND THE PRODUCT LIFECYCLE

3.1 Introduction

The engineering domain is part of the lifecycle of an electrical, electronic or electro-mechanical product focusing on the phases from requirement of a new product until it gets shipped to the customers. In a wider sense it touches also a product's life until its disposal and recycling, as shown in figure 1, in the way, that once a product got shipped to a customer or is available on the market, maintenance and repair has to be done.

Today's companies are driving their product development with sophisticated processes and applications, to reduce time-to-market of their products. Therefore in every phase of the product

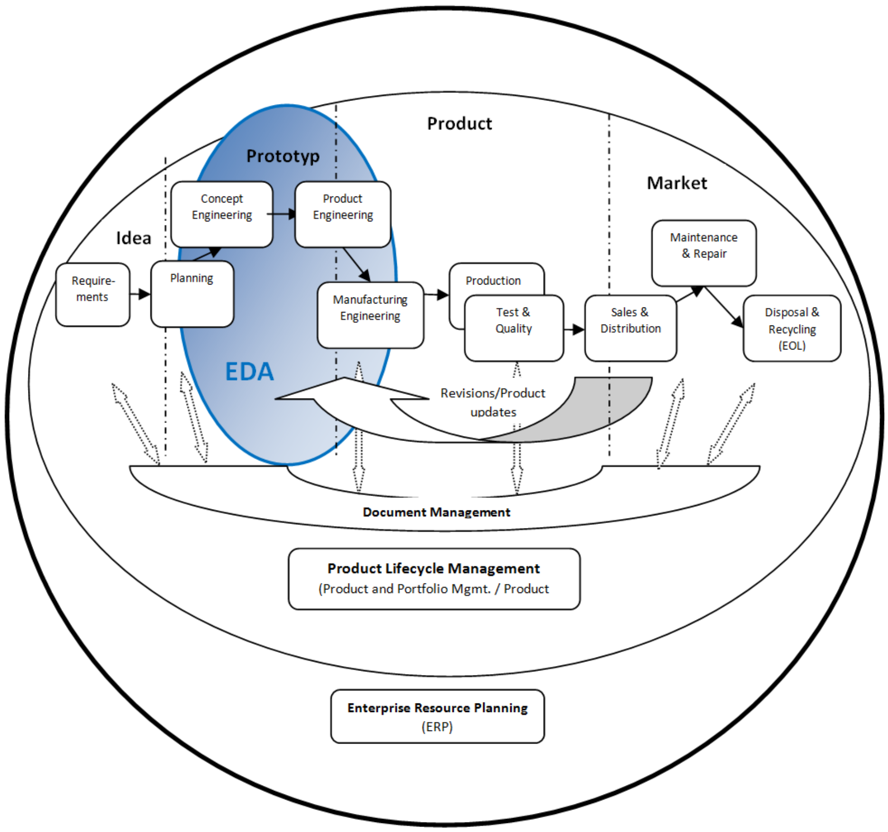

Figure 2: Simplified Product Lifecycle

lifecycle, software applications are used and data is exchanged between these applications.

3.2 Product Lifecycle

For a better understanding the typical lifecycle of a product is briefly described below including the flow from the idea for a product until its end of life. The duration of a product's lifecycle can vary from a few months to a year (e.g. cell phone, mp3-player), over three to ten years (e.g. computer, automotives, medical devices, software) to 50 and more years (e.g. airplanes, defense weapons). The flow usually starts by entering basic data for a product into an ERP system, so that it can be looked at by various roles in a company.

3.2.1 Enterprise resource planning

Enterprise resource planning (ERP) is an enterprise-wide information system designed to coordinate all resources, information, and activities needed to complete business processes such as order, resource planning, order fulfillment, billing, and to support separate applications like supply chain management (SCM) and customer relationship management (CRM). An ERP system usually can be made interoperable with a PLM or a PDM system.

3.2.2 Product lifecycle management

Product lifecycle management (PLM) is the process of managing the entire lifecycle of a product from its conception, through design and manufacturing to service and disposal. PLM integrates people, data, processes and business systems and provides a product information backbone for companies and their extended enterprise.

PLM systems cover four primary areas: product and portfolio management (PPM), product design, manufacturing process management (MPM) and product data management (PDM).

PDM systems maintain a single copy of the product master data in a repository, often called secure vault; the data are then distributed to those departments requiring them. PDM systems maintain a product history and configuration management of the data. PDM and PLM systems are used in the engineering and design communities. After a project was entered in ERP it can be populated into PLM/PDM, which in essence means the metadata will be transferred into PLM.

3.2.3 Document Management

Sometimes engineering companies also use Document Management Systems (DMS) to handle all aspects of documentation. Those DMS may be integrated with the other systems (ERP, PDM or PLM). If no such system is in place, documentation will be usually stored in the PDM system.

3.2.4 Electronic Design Automation (EDA)

Electronic design automation is the integral part of the electronic product lifecycle, where products get defined, prototyped, developed, tested and then send to manufacturing to be build. This is the "world" of Computer Aided Design (CAD), which is the use of computer technology to aid in the design and particularly the drafting (technical drawing and engineering drawing) of a part or product, including entire buildings. CAD is sub-divided into Mechanical Computer Aided Design (mCAD), and Electronic and Electrical Computer Aided Design (eCAD).

It is important to understand that EDA represents design processes or flows, as illustrated by figure 2. These flows take a design from an engineer's concept, to a logical representation, to physicalization (physical design) - the data needed to fabricate a chip, a circuit board, or even a complete system.

3.2.5 Product Definition

The product definition process starts with requirements coming from a potential market by monitoring that market, doing customer surveys, collect requirements on exhibitions or even define a new market.

3.2.6 Concept engineering / Prototype

Once the new product has been approved by management, a design concept will be developed and a prototype will be designed. In this phase EDA software will be used, where mechanical engineering is developing the 3D mechanical parts of the product, and electrical engineering is developing the electronic portions, mainly the printed circuit boards (PCB).

3.2.7 Product Engineering

Once the prototype has been built successfully and management has approved to go further, electrical and mechanical engineers develop (sometimes in several iterations and most often in a concurrent way) the final 3D mechanical pieces, printed circuit boards, cabling etc. Once both design communities have finished the development phase, a defined set of data will be passed to manufacturing for further processing.

3.2.8 Current Engineering Archiving/Preservation

Today, archiving of product data is performed in several systems throughout a company. While marketing documents often reside in an ERP and in requirements management systems, production and commercial data in ERP, design and engineering data is mainly preserved by facilitating PDM or PLM systems. Those systems usually keep the specifications, design files, component libraries, simulation models, simulation results and those production data generated from the designs (e.g. Gerber files). Checking out a complete set of data for a particular product, means to access all these different systems and to research and discover the correct data and versions. Beside this challenge some product data is usually redundant between those software solutions, because portions of that data will be used for different purposes in all those systems (example: a marketing requirement document is created and maintained in a requirements management system, but is also needed in PDM/PLM to create functional specifications and in ERP to allow purchasers to check pricing and availability of electronic components).

3.2.9 eCAD/mCAD Collaboration

Nowadays efforts have been made to establish cooperation capabilities between the different design domains. As an example the collaboration between mechanical an electronic design can be mentioned here. Due to the different systems on the market today, these domains have only limited means for exchanging information. Thus, it is difficult to ensure that a PCB developed in the eCAD system fits into the enclosure which is developed by the mCAD system. In a ProSTEP project for this purpose a new standard has been developed which allows the exchange of information between the domains which enable the engineers to collaborate across the domain boundaries [15].

4. PRODUCT DATA

In this section we give some more insight into the data which are created during the design process. By looking at it from the perspective of various views we identify main characteristics of design data.

4.1 Data created during the design process

During the **Product Definition**, data is captured often by normal document tools like MS Office or OpenOffice, but often this information is also kept in proprietary formats, if commercial requirement management systems are used (like IBM Telelogic Doors or Rational RequisitePro).

During **Concept and Product Engineering**, a number of phases have to be passed which all deal with their different design objects:

- **Library Creation:** Logical and physical design require a component library, which contains technical data for a component and graphical symbols as representations shown later in the design applications. Libraries are either in a proprietary format or kept in a database. They are made up of metadata and graphic data.
- **Schematic capture**: The output is an electrical representation of a design, often called the netlist, which is a list of components and their connections. Most common – beside proprietary formats for a netlist is EDIF [5]
- **Component Placement**: The output is a list of components and their x,y coordinates. The output can be simple ASCII, but is most often proprietary and binary.
- **Routing**: During routing, the physical connections between components are created based on the logical interconnections of the schematic capture. The output is in almost all cases proprietary binary and contains beside the layout graphic of the component (footprint) their x,y coordinates and the side of the pcb, where it is placed; plus the x,y coordinates of each interconnection segment, its width, on which layer of a pcb it is placed etc.
- **Analysis and simulation**: For the logical design, timing, fault, and circuit simulation can be performed. The simulators use the logical design information and respective simulation models. The outputs are e,g, wave form and timing diagrams. For the physical design thermal, signal integrity and electromagnetic compatibility analysis can be performed. They use the physical design information and respective simulation models. The outputs are waveforms and diagrams. In addition testability and manufacturability can be simulated.

4.2 Dimensions of product data

Objects which are used in engineering environments are in many respects similar to objects used in other domains. However, if we look at them under the perspective of a number of views which include the context in which they have been generated, the way they are supposed to be processed further and their relationships amongst each other, they are in many ways different from objects which are dealt with in many classical preservation scenarios.

Let's start with an (exemplary) list of aspects which can be used as dimensions – and product data tends to position itself on the left side in each single line of the following table 1 where explanations are given in the specified subsections:

Based on formal model, supposed to be processed by further tools	⇔ 4.3	Informal model, normally not supposed to be processed in an automated way
Related to complex, manifold tool suite	⇔ 4.3	Related to single tool for manipulation and viewing
Reusable	⇔ 4.3	viewable
Complex relationships between design objects – no holistic representation	⇔ 4.4	Single, standalone objects
Various kinds of metadata	⇔ 4.5	Little metadata
Process based developments	⇔ 4.6	Objects are not related to creation process

Table 1: Dimensions of product data

4.3 CAD data and tools

Data which is preserved in classical applications for digital preservation tend to be preserved for viewing it later in time. Thus, even if there is some software necessary (like a pdf reader, etc.), this software often does only display the object but not necessarily process it.

CAD objects in general are highly structured and interrelated. The purpose of CAD data is the representation of the result of a specific step in the design process which will be further processed in the next design steps in various ways. This is done by means of quite complex design tools like analysis tools (e.g. logic, timing, temperature, collision, etc. analysis or simulation) or generation tools (e.g. layout synthesis tools, routing and placement tools, etc.). These tools generate additional data, either to present their analysis results or to create the basic data for the next design phase. Thus, CAD data is very often related to the generator tool which created it and it has to fit to the tools which will process them in the next stage.

4.4 Partial views

An important aspect is that in most cases no single design object represents the final product in total. A holistic view of all aspects of the object can only be achieved by a combination of the design objects which each only provide a partial view, e.g. netlist, layout, behaviour, geometry of the complete design. Thus, to reuse the design later in time, it is not sufficient to store a few specific design objects, because that could mean that some information is missing to further develop it in all aspects.

4.5 Metadata

In design a multitude of metadata is used, and quite often, it is difficult to differentiate between data and metadata. For example, design objects are usually composed of other design objects. The composition is done by means of an instantiation of the component in the context of the composite design object. This component instance is only a light object (e.g., only with information about its relative position and orientation) which refers to its "master" component object [5]. This master is a schema for all component instances of the same type, i.e. it provides the metadata to specify these component instances. The master contains information like

- parametric information (e.g. dimension, electrical properties, etc.)
- simulation model(s) to allow the integration of the component into simulations of the complete design
- interface information, e.g. to allow the placing of the component on a board and to connect it with other components

Additional metadata is used to distinguish different versions of the same design object, to specify reasons for decisions or to annotate by which tool under which operating system a design object was created. Thus, there exists metadata to capture the context of the design object.

Another kind of metadata is used to standardize the used properties: In international product standards (ISO 13584-42 [10], ISO 13584-511 [11], IEC 61360 [7], etc.) and by industry consortia (like eClass [4], UNSPSC [18], etc.) ontologies or classifications have been developed which allow to associate products to specific product classes and to annotate properties with a specific meaning defined in external ontologies. Thus, these ontologies are another form of metadata for CAD objects.

4.6 Process orientation

As mentioned above, design objects are usually created by complex processes, and in most cases they represent only a specific view of the complete design. This relates to the various views which are required to fully describe complex products, however, it also relates to the process by which this product has been designed. The result of a process step is a new design object which will be further processed in the next design steps. Thus, a design object also represents part of the development process, and the many design objects provide a trace of the activities from idea to product. This is particularly true in today's PDM and PLM systems, which provide complex version and configuration control mechanisms enabling tracking of the design process including the capture of decisions, their rationale, who has authorized them, etc.

5. CHALLENGES

Based on the description of the engineering process and the resulting product data in the previous chapters, a number of challenges can be derived that have to be addressed on the way to a sustainable approach to digital long term preservation of engineering data. The enumeration of challenges presented below can be categorized as follows:

- Strategic challenges are linked to the economic benefits of long term digital preservation. Managers will only support the introduction of DP if there is sufficient return on that investment (ROI).
- Technical challenges are related to the underlying concepts, architecture and implementation of DP systems. They often are the other side of a strategic or organisational coin.

Challenge 1: Evaluation of economic implications

As described in section 2, by providing sophisticated DP systems several benefits are achieved. For example, current insufficient and ineffective archiving methods (paper archiving, archiving by keeping old hardware and software) are replaced. Another benefit is the increase of opportunities to reuse existing designs (without long-term preservation of designs reuse is not very likely).

There is no doubt that implementing new DP functionality, which has to be done in addition to the existing design process functions, will create additional costs, as discussed by the Dutch National Archives [2]:

- Product data need to be stored on reliable storage media. The more data exists on media the more data needs to be secured from risks, such as hardware failures or unauthorized access which results in higher cost of **hardware and its administration**.
- Additionally in house or third party **software extensions** are necessary to perform DP.
- The long-term archiving process requires additional activities which might lead to higher **cognitive load**.
- The additional DP functions might lead to additional **personnel costs**.
- Employees need to be **trained on DP skills**.

DP will have a chance in business activities only if it can be shown that the benefits outweigh these costs. Thus, it is necessary to perform cost analysis of use cases to expose the benefits which can be achieved with a next generation DP system. These investigations need to take into account the purpose of the archival (legal reasons, reuse), the industry branch and also the kind of product to be developed, including the longevity of the product. In addition, the result of such an analysis will be dependent on the technical features which can be assumed to be available in a DP system. We need to get qualitative results (which additional opportunities will arise from better DP systems) and quantitative results calculating cost reductions and benefits due to new functionalities and opportunities. However, calculations of the return on investment (ROI) are depending on the motivation of a company for digital preservation, as it makes a huge difference whether a company is doing DP for certification, legal, reuse or “support in production operations” purposes. A more detailed discussion of ROI would therefore go beyond the scope of this paper.

A company that has decided to go for a long term archiving of its engineering data has to address a number of organizational and technical issues. These issues will be discussed in the following as technical challenges. For supporting companies in taking the right decisions, best practices have to be developed to guide the use of DP in engineering projects.

Challenge 2: DP integration into processes

As we described in the previous chapters, the engineering activities are embedded in processes which are performed by use of a number of tools and systems. In particular PLM systems control the development process and ensure that the engineering activities follow predefined rules.

It has to be investigated how DP affects the existing processes. That means that the DP functions must be integrated into the existing engineering processes. One idea to be considered is to identify overlapping areas or similarities with existing functions of PLM systems, e.g. whether DP functionality can be realized as extensions or additions of existing PLM functions. Another aspect is the introduction of new organisational functions and new roles of personnel which might be necessary for operating a DP system. There are activities required during the design process (ingestion) and during the preservation phase (maintenance of preserved data).

One basic source of information should be the results which have been achieved in other preservation activities [6]. Domain specific process variations in long-term archiving processes might exist so that process integration has to allow for adaption to different domains and different company cultures.

Challenge 3: DP integration into engineering systems

In addition to the organisational issues, also the system specific issues have to be addressed. This requires first a clear specification of the long term archiving activities and the development of concepts and architectures to integrate them into the existing engineering systems consisting of ERP, PLM and CAD tools.

The DP process certainly includes important activities like ingesting and accessing product data [1]. Ingestion is the process of archiving data and associated metadata into the archiving system whereas accessing means to find, view and reuse archived data. High data quality during ingestion is important because there is a trade-off between the effort for achieving high quality ingestion and the effort of retrieval such ingested data: The more effort is spent during ingestion the less effort is needed during retrieval and vice versa. But in order to lower the costs for archiving the long-term archiving process needs to be **as effortless and transparent** as possible. Therefore, a possible solution might include the reduction of efforts during ingestion by the means of automatic extraction of metadata and validation of manually entered metadata.

Challenge 4: Standards for tool interoperability

Currently, many tool providers use their own proprietary formats and data models to capture their design knowledge. This is an obstacle for interoperability between tools of different vendors. In the same way, the component library information which are provided by component manufacturers and which are supposed to be integrated into the library of a design system are not really following standards so that their integration into the system requires translation, reformatting and often manual intervention.

On the other hand, there exist a few standards which are used in some areas, e.g. STEP [9] for exchange of geometry information between CAD tools or the IEC 61360 dictionary of electrical components [7] and the Japanese ECALS dictionary [3]. To strengthen the interoperability among design tools and design steps and to facilitate the reuse of preserved design data after a long archiving period, the use of standards is of high importance and should be encouraged wherever possible.

Thus it is a challenge to examine areas where standards are missing since a variety of artefacts can be subject to standardization including service interfaces, data and metadata. One example is the eCAD-mCAD collaboration project of ProSTEP [9] where a model has been developed which allows the exchange of CAD objects that are relevant both for eCAD and mCAD engineers to enable them to communicate across the different domains.

In current PLM processes, also tools which work on their proprietary formats are integrated by techniques of the PLM systems. It has to be examined if and how far these techniques can be exploited also for digital preservation in the long term.

Challenge 5: System environment modeling and preservation

The long-term ingest activity consists of the preservation of relevant artifacts to meet both legal and economic requirements. But the complete product state does not only consist of product data, database contents and documentation, but also of hardware and software by which these data have been created and manipulated. Note, that the term software covers all relevant software like embedded software, database software, application software and operating systems that were used while working with product data that is to be archived. Modeling and preserving this context state sounds easier than it is because figuring out which artifacts to archive is regarded as a hard technical challenge. For example, product data is developed by using resources that are distributed over the internet by using cloud computing services. Archiving the complete server, service and software environment turns out to be a severe problem that might be even not solvable.

Also product data is processed by tools that are subject to technological development. If tools evolve and are not able to process archived data then this data is of only short usefulness and has a high obsolescence rate so that **the life cycle of tools and data differ**. The challenge is to guarantee that product data that originates from an obsolete platform will still be exploitable, even for tools that access the product a long time after the product data were created. This core technical challenge is addressed in the SHAMAN Project [15, 18]. To solve this challenge and prepare product data so that it remains useable in the future at least four strategies exists that need to be evaluated if they are adequate for product data:

- During **migration** the data is transformed from one format into a more recent or another format. If migration is used as transformation then consistency, authenticity, integrity and references to other artefacts needs to be guaranteed. Also, if ingested metadata is specified according to a specific schema then it is possible that also metadata needs to be migrated to a new schema version [12]. Solutions are proposed by existing projects like KIM [13] and Digital Engineering Archives [16]. These projects propose a registry for engineering data formats which supports migration of product data for long-term access and reuse.
- During **emulation** the data is not transformed. Emulation rather creates virtual run time environments (CPU or operating system) so that unmodified product data can be used [12].
- During **encapsulation** product data is retained in the original format but contains instructions on how the original CAD data should be interpreted via metadata [12].
- The **multivalent engine technology** preserves the ability to manipulate the original data format without the original software [14].

Challenge 6: Identification of relevant data

As we have seen previously, during the design processes a huge amount of different objects are created. Preserving them all may easily result in a confusing data collection where relevant data can not be distinguished form irrelevant data. Long-term archiving is pointless without the possibility to use and interpret existing data in a sensible way. Selecting the amount and type of data to archive is a challenge. Also, if the effort for finding data is high it will result in a lack of acceptance of archived data. In order to enable a searchable and reusable data archive it is necessary to spend more effort during initial ingestion by specifying metadata that helps to navigate through archived data in the future.

Therefore, as a conceptual challenge, it needs to be figured out which aspects (environment, process, etc.) need to be described by data and metadata and which retrieval methods are best suited. One kind of data to be preserved are the design objects which describe the design and the product. If the process itself needs to be tracked, then a number of additional information has to be preserved like (1) change requests, (2) resulting modifications, (3) persons who have performed the modification or have approved design changes, (4) information about the tools which have been used for the generation of a design object. This data will also allow to track collaboration of designers (e.g. cross domain, cross organizational). A further kind of (meta-) data which possibly needs to be preserved are elements of external ontologies (e.g. IEC 61360-4) defining the meaning of properties, classes and concepts which are used for the component, design and process description.

A related aspect to consider is the difference between automatically generated objects (e.g. results of simulations steps) versus interactively-created objects where humans have intervened. If the quality of the preservation of the computing context (see challenge 5) is good enough, then automatically generated objects might not be archived.

The time for which a design object has to be preserved might vary depending on legal requirements or the type of object. If the specified preservation duration of a design object has elapsed or if product data is not needed anymore for economic reasons relevant objects can be deleted. Thus, the identification when and which **artifacts are disposable** is another challenge. In addition, rules for deletion have to be defined so that a deletion process might be automated with minimized human interaction.

Challenge 7: Distributed, secured long-term archives

In today's business, companies interact with external partners and other companies because complex projects cannot be performed by one company alone and because for many projects specific know how is required which is "bought in" from partner companies and which allows the integrating of best of class components. Engineers that work for those corporations and joint ventures often develop products in geographically distributed environments. In such distributed environments data also needs to be archived for the same legal or economic reasons as for non distributed environments. Therefore, as a conceptual and technical challenge, we need to think about processes and interfaces for **distributed long-term archiving** that provide means of accessing or even importing of archived data.

In these environments specific privacy concerns have to be considered. In particular, in case of **cross organization collaboration processes**, partners only share knowledge which is required to work on a specific project. Companies do not disclose all information about their components or their know how to their collaboration partners. If DP is necessary in such a situation, companies will not disclose more information for long term preservation as they would disclose in the collaboration process itself. There are at least to ways to solve this problem: (1) distributed DP where each partner is responsible for his objects, or (2) central DP where the partners require means which prevent a disclosure of their data unless a specific event or case occurs. For example, it needs to be taken into account that after a company bankruptcy access to archived data of that company is ensured. It is both a technical and conceptual challenge to reduce such **privacy concerns and third party privacy seals** so that protection from unauthorized access to the whole or non legal parts of the data is guaranteed.

6. CONCLUSION

The previous sections described the general design and engineering scenario, processes and data. There are significant economic and legal reasons for engineering companies to invest in digital preservations technologies and research.

The range of challenges for preservation technologies to support such companies has many dimensions, ranging from strategic analysis to practical technical issues.

In our opinion, due to the complexity of the topic it is unlikely that in the foreseeable future a comprehensive practical system will emerge, addressing the preservation needs of all thinkable engineering environments and scenarios, while solving all of the presented challenges. In addition we assume, that there are hard practical limits regarding the archival of engineering environments, based on the distributed nature and the increasing amount of cloud computing resources in the engineering process.

However, there is a multitude of opportunities to address practical preservation problems within the scope of individual scenarios and engineering domains. It is now important to identify practical niches of this kind to drive forward research and product development for solving the existing preservation problems of engineers, and to evaluate how these detail solutions can all come together for providing a more general and sustainable approach to the general problem.

In the research project SHAMAN [17], we are examining how it is possible to apply and adapt digital preservation approaches to solve problems in the domain of electrical engineering and collaboration. To do so, we examine the relation between the evolution of design environments and the evolution of archive systems and design new preservation-aware collaborative tools for design and engineering.

7. ACKNOWLEDGMENTS

This paper is supported by the European Union in the 7th Framework within the IP SHAMAN.

8. REFERENCES

[1] Consultative Committee for Space Data Systems: Reference Model for an Open Archival Information System (OAIS). January 2002. http://public.ccsds.org/publications/archive/650x0b1.pdf

[2] Dutch National Archives, 2005. Costs of Digital Preservation. Available from: http://www.digitaleduurzaamheid.nl/bibliotheek/docs/CoDPv1.pdf

[3] ECALS/JEITA: ECALS dictionary of electronic components, http://www3.elisasp.net/ecals/

[4] eClass: International standard for the classification and description of products and services, http://www.eclass.de/

[5] Electronic Design Interchange Format; www.eia.org

[6] Higgins, Sarah, 2008 "The DCC Life Cycle Curation Model." International Journal of Digital Curation, 1:3 (2008).

[7] IEC 61360 database: Standard data element types with associated classification scheme for electric components – database with IEC reference collection of standard data element types, component classes and terms, http://std.iec.ch/iec61360

[8] ISO 9000; ISO – International Organization for Standardization; http://www.iso.org

[9] ISO 10303: STEP – Standard for product data representation and exchange. http://www.TC184-sc4.org

[10] ISO 13584-42: Industrial automation systems and integration -- Parts library -- Part 42: Description methodology: Methodology for structuring part families. 1998

[11] ISO 13584-511: Industrial automation systems and integration -- Parts library -- Part 511: Mechanical systems and components for general use -- Reference dictionary for fasteners. 2006

[12] Lee, K.H., Slattery, O., Lu, R., Tang, X. and McCrary, 2002. The State of the Art and Practice in Digital Preservation. Journal of Research-National Institute of Standards and Technology, 107(1), pp. 93-106.

[13] Patel, M., Ball, A., Ding L., Strategies for the Curation of CAD Engineering Models. In: 4th International Digital Curation Conference. 1–3 December 2008, Edinburgh.

[14] Phelps, T.,Watry, P.: A No-Compromises Architecture for Digital Document Preservation. In Procs. Research and Advanced Technology for Digital Libraries, 9th European Conference (2005), pp. 266-277.

[15] ProSTEP/IVIP: eCAD/mCAD Collaboration, ProSTEP/IVIP Recommendation, 2008, http://www.prostep.org

[16] Regli, W., Michael J. Grauer, Iris K. Howley, Joseph B. Kopena. Towards a Format Registry for Engineering Data, ASME International Design Engineering Technical Conferences & Computers and Information in Engineering Conference, 4-7 September, 2007, Las Vegas, NV.

[17] SHAMAN - Sustaining Heritage Access through Multivalent ArchiviNg, http://shaman-ip.eu/shaman

[18] UNSPSC: United Nations Standard Products and Services Code, http://www.unspsc.org/

[19] U.S. Department of Health and Human Services - Food and Drug Administration, Guidance for Industry, Part 11, Electronic Records; Electronic Signatures - Scope and Application; http://www.fda.gov/cder/guidance/5667fnl.htm

[20] Watry, Paul, 2007. Digital Preservation Theory and Application: Transcontinental Persistent Archives Testbed Activity. Int. Journal of Digital Curation, Vol 2, No 2.