Computers in Industry 65 (2014) 24–36

Contents lists available at ScienceDirect

Computers in Industry

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

Preservation-awareness in collaborative engineering

Dominic Heutelbeck *, Jens Grabarske

Forschungsinstitut für Telekommunikation – FTK e.V., Martin-Schmeißer-Weg 4, 44227 Dortmund, Germany

ARTICLE INFO

Article history:
Received 10 September 2012
Received in revised form 17 May 2013
Accepted 10 July 2013
Available online 5 September 2013

Keywords:
Digital preservation
Product life cycle
Collaboration
CSCW
Social networks
Workflow capturing

ABSTRACT

In design and engineering, it is important to preserve more than the actual documents making up the product data. For knowledge-intensive industries it is of critical importance to also preserve the *soft knowledge* of the overall process within the product life cycle. The idea is not only to preserve the designs for the future, but also the knowledge about processes, decision making, and people. In order to preserve this knowledge, it is necessary to captured it at content creation time, a process currently mostly independent from the preservation process. This paper discusses how to make applications in content creation (e.g., in design and engineering) preservation-aware by using the OpenConjurer approach and framework.

1. Introduction

For innovation driven industries like design and engineering, digital preservation is an important challenge. Legal regulations and contractual requirements often make the preservation of product data mandatory for decades. Product data is all data generated during the product life cycle, e.g., text, presentations, simulation data, and designs. The reuse of product data and knowledge in a rapidly evolving technological environment is an important a goal presenting difficult problems. The motivation for digital preservation in design and engineering is discussed in detail by Heutelbeck et al. [1,2]. A key motivation is to preserve and reuse the knowledge about the actual design and decision making process. This includes documentation about who did what with the product data. In this paper, the 'who' part of this information is called *social context* and not only consists of information about the unique identity of individuals, but also about their role and position within the social network of the people and organizations involved in the creation of the product data. In addition, the 'what' part of this documentation is called *collaboration context* and covers operations on the actual data as well as communication and collaboration surrounding the data, such as decision making, i.e., how something was done or decided. This includes the processes as part of the overall collaboration-context.

As an example consider the following scenario. A company plans to release a second generation of an existing product. The original product developed a failure pattern during maintenance of the product in use. It turns out, that a component was under dimensioned for the actual usage patterns of the product. The engineers of the company try to investigate, what the reasons for this where and if this was a singular decision of one person or if the issue had been discussed within the group of product designers. So the question is, if there where group discussions resulting in an explicit decision leading to the dimensioning of the component failing regularly. The engineers plan to use this insight for the redesign of the second generation of the product. The original product has already completed the pre-production part of its product life cycle, and is already archived in a preservation system. So the engineers consult the preservation system and search for design decisions captured for the failing component.

More generally the following questions came up repeatedly during requirements analysis and discussions with industry users and experts:

1. Forensics and analysis of the past
   (a) Reasons for commercial or technical failure?
   (b) Which decision or negotiation process led to a fault?
2. Knowledge reuse
   (a) How can we improve our products?
   (b) How can we improve our processes?
3. Design reuse
   (a) How can we enable a complete reuse of original designs in a new design?

* Corresponding author.
E-mail addresses: dheutelbeck@ftk.de (D. Heutelbeck), jgrabarske@ftk.de (J. Grabarske).

0166-3615/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.compind.2013.07.002

(b) How can we reuse functional parameters for reengineering or maintenance?

Especially for questions 1a, 1b, and 2b the preserved knowledge about the processes and the social context of the content creation is important to both find the appropriate content as well as for achieving the goals of the users.

A variety of approaches for product life cycle management exist, without limiting the general applicability of the approach, this paper uses a product life cycle that distinguishes between well-defined phases as an example. In such a scenario, it is reasonable to perform archival of a set of data into a preservation system at critical points in the product life cycle. Such dedicated points in time are often the transitions between different phases of the product life cycle. Typical points would be the transition between the ideation phase and the development phase, or the transition between development and production of a product. The phase before the archival can be arbitrary long. Thus, it is difficult to manually recreate the knowledge about social and collaboration context at archival time for reasons of human memory, availability, and cost. This paper discusses an approach to enable the tools used during the creation of the product data to become aware of the social and collaboration context, to enable them to express themselves in the terms of these contexts, and to capture the knowledge for preservation and reuse semi-automatically.

The paper is structured as follows. First the requirements for making applications preservation-aware are discussed, and then related work is presented. The next sections discuss the Open-Conjurer approach of providing a flexible ontology-based vocabulary and infrastructure. This is followed by the introduction of a demonstration application which is used to illustrate how to implement a preservation aware collaborative application in practice. Then it is discussed how to integrate the preservation aware applications with content repositories, product life cycle management (PLM) systems, and preservation systems. The paper concludes with findings from demonstration and evaluation, a discussion and an outlook to future work.

2. Hypothesis and problem statement

This paper is based on the hypothesis, that in design and engineering, it is important to capture the design process.

It is assumed, that most collaboration tools in use are not able to capture much of the design process. The problem addressed by this paper is the question on how to enable the tools used during the creation of the product data to become aware of the social and collaboration context, to enable them to express themselves in the terms of these contexts, and to capture the knowledge for preservation and reuse semi-automatically.

The first sub problem in order to capture such processes and social context is to specify the matching vocabulary describing the social context as well as the collaboration context. As individual enterprises differ significantly regarding their structure and corporate culture, the vocabulary needs to be customizable to match the real world context of use. In addition to the vocabulary it is also necessary to operationalize the management of an inter-organizational social graph by providing an infrastructure and APIs.

The second sub problem is that in practice, product data is stored across a very heterogeneous population of document management systems, file shares, and PLM systems. In order to enable the capturing of contextual metadata, product data repositories are necessary that offer interfaces for annotating product data with custom metadata. For integration with the overall workflows such as decision making or archival the repositories need to be linked to legacy PLM systems. For archival, the repositories need to provide interfaces for collecting the product data and collected context metadata.

3. Related work

In digital preservation, the OAIS reference model [3] is commonly adopted and illustrates the common understanding of preservation systems and workflows adopted by the digital preservation community. In OAIS, the tools used for content creation are mostly unaware of the archival system. Data from a producer is collected into a so-called Submission Information Package (SIP) which is ingested into the archival system. Here, it is transformed into a so-called Archival Information Package (AIP) and is preserved within the archive. In essence, this means that the actual data from the producer is created independently from and unaware of the preservation environment. Usually during the creation of the SIP, metadata for the data is cleaned up and added, either automatically or with manual intervention. The result of this view of preservation being completely decoupled from the actual data creation results in scenarios, where metadata, which is readily available during data creation, is lost before reaching the creation of the AIP. For example, change logs and user information are available in MS Office documents, but in practice digital libraries often only archive a PDF version of the source material. In the conversion process usually most of the metadata is lost. While it is certainly possible to archive the original documents, the loose coupling of the preservation systems with the content creation environment and the lack of awareness of the content creation tools for requirements from preservation result in this common loss of important data.

The concept of gathering information about what happened in a collaborative environment is not new, and is called activity log, event history, or elephant's brain. From a software engineering perspective this can be considered a design pattern and is described in detail by Schümmer and Lukosch [4]. The intent of this pattern is to: "Store information about users' activities in a log to provide a history of their activities and the artifacts' evolution." This pattern is regularly applied in various collaboration environments, and is especially prominent in software development environments. A system keeping an activity log is already capturing a basic social context (user names) and collaboration context (what has been done).

However, in practice the types of activities to be logged are hard-wired, and the vocabulary is not customizable. Activity logs are typically geared towards use within productive environments and not towards digital preservation.

Within digital preservation logging usually happens within the preservation systems in order to keep track of the preservation process and changes to preserved documents. In this context this type of information is called *provenance* as discussed in [3].

It is important to distinguish these two types of logging information clearly. Activity logs capture information about what happened with an artifact during its creation and active use, which means before the ingestion of the artifact into a preservation system. Currently this type of information is mostly lost, e.g., through publishing processes or during ingest. After ingest the logging mechanism of the preservation system takes over and logs what happens to the artifact within the system. This usually includes preservation activities such as migration or replication.

This paper aims at capturing activity logs in a well-defined open scheme, which survives the ingest process into a preservation system and is able to be preserved and indexed.

In order to discuss social graphs and models for social context we have to distinguish between the sociological models and data models for social relations.