Digital Preservation: New Challenges for Software Maintenance

Sven Euteneuer, Daniel Draws
SQS Software Quality Systems AG
Stollwerckstraße 11, 51149 Cologne, Germany
Email: sven.euteneuer|daniel.draws@sqs.com

Abstract: Digital preservation ensures continued access to digital information over time (e.g., 5, 15, 25 or 50 years). The long-term planning horizon of preserving relevant parts of a business processes raises a number of new research challenges for software design, development and maintenance (cf. [1]).
In this paper we describe the challenges caused by the preservation of software and give a short overview of the TIMBUS research project that aims to address these challenges.

1 Introduction

Business processes are an orchestration of different services. Many of these services are operated by different service providers. Processes are volatile in terms of services disappearing and fundamental changes in technology. Processes of vital interest in a remote future need to be preserved. Their functionality, usability, integrity and authenticity need to be guaranteed. Changes of technology (e.g., new formats or standards), changes of the environment (e.g., new legal obligations) and the disappearance of services need to be addressed for ensuring the usability of the process and its software in the long run.
Current maintenance approaches for software are dealing with online (and active) systems. Changes in the environment, technology, services and data potentially have immediate effects on the system functionality. For preserved software stored off-line in an archive these changes can remain concealed for years. New approaches for active management and maintenance of preserved information systems are required. The risk of irreconcilable inconsistency of the software system caused by technological obsolescence or unavailability of essential services and software components need to be addressed.
Preservation of software systems raise new challenges for all stages of the software life cycle. Starting from the design phase, well-documented and clearly designed structures of software reduce the dependencies on actual implementation technology. Other examples are requirements on the software quality raised by the fact that source code should be understood in the future.
Maintenance over the long term can include the replacement of obsolete technologies, such as runtime environments, operating systems, networking infrastructures, etc.
The TIMBUS research project addresses these and other research issues, amongst which are authenticity of the preserved information, legal, contractual and regulatory issues and risk management with respect to business process preservation.

2 Digital Preservation

The preservation of digital artefacts was identified as a serious issue and is being addressed by different research efforts over the course of the last years. Digital Preservation -ensuring continued access to digital assets- is traditionally focused on data-centric information. A number of research initiatives have addressed preservation of digital objects, such as office data, video or more complex data structures such as databases. An overview about relevant research projects within the ICT program of the European Commission is given in [2].
The data centric view considers the generation of data from a bitstream. This is addressed by the definition of representation information (c.f. [3]). A powerful representation information is the software that generated the data. Concerning the preservation of information for decades it is obvious that this software would need to be maintained for a long time. The difference to the established maintenance approaches is the fact that archived software stays unchanged for a long period. The development process often has to be re-instantiated in a completely different – and ex-ante unknown – environment. Exclusively preserving and provisioning source artefacts would comply with such a data centric preservation approach.
Unfortunately, the data centric view leaves out the relevance of the data's context. The context transports the semantics required to transform the captured data back into information. It also ignores important aspects of execution such as processing, analysing, transforming, rendering and interpreting.
Workflows and business processes (as well as software development processes) often rely on a variety of sub-processes and distributed, loosely orchestrated and service oriented architectures. Unavailability of data, services or context can lead to loss of information resulting in detrimental effects on the businesses they enable. Digital preservation of processes and services, as it is addressed by TIMBUS, is a novel approach, because it addresses the preservation of information in a holistic manner.

.

3 Holistic Preservation

Successful digital preservation of whole processes requires capturing sufficient detail of a process and its context to be able to re-enable its original behaviour at a future date, involving potentially different participating parties, different enabling technologies, different system components (hardware and software), changed services by different service providers or differences in other aspects of the context of the business process. This implies that there exists a set of activities, processes and tools that ensure continued access to services and software necessary to produce the context within which information can be accessed, properly rendered, validated and transformed into knowledge. Digital preservation of business processes and services therefore requires that this set can be preserved.

Traditional digital preservation approaches have a focus on preserving digital objects and their technical context [5]. However, from a TIMBUS perspective, traditional methods of digital preservation do have a too restricted perspective on context that they take into account. Assuming that a traditional method of digital preservation provides the ability to preserve system components and entire systems in addition to digital artefacts, often, it would not be clear at a given time, which information would be required later to restore the system. Things that constitute a clear boundary of technical feasibility today can be a parameter not accounted for the future.

The approach developed in TIMBUS is expected to identify all relevant artefacts generated and used in a business process. TIMBUS investigates to identify and resolve dependencies under given constraints to ensure re-executability of processes in the distant future. For the long term preservation of software systems with a holistic view all artifacts used in a development and deployment process have to be considered [4]. Besides the technical preservation of the software system, the context of business processes in question may be influenced by a variety of changes in the environment (such as change of legal aspects over time or the change of tacit knowledge). To give an example: Notations like UML are accepted and well understood today. General concepts and most notations can be used by developers without looking into the standard itself. But there is an uncertainty that this will be the same in about 50 years or more. The preservation process proposed in the TIMBUS project consists of three stages of digital preservation:

Expediency of digital preservation effort – establishing the risk of not preserving and the feasibility of digitally preserving business processes.

Execution of digital preservation process – performing the digital preservation of business processes.

Exhumation of digitally preserved assets – re-running a digitally preserved business process.

4 Summary and Outlook

TIMBUS is an extended and more curation-oriented approach for preserving business process. Adopted to software maintenance this implies a more holistic view on a software project also considering other collaterals besides source code. Expected outcomes of the project are amongst others:

- A risk model for business processes with respect to digital preservation.
- A context model for business processes.
- Preservation processes required to preserve business processes in the long term.
- Guidelines for ``digital preservation-aware'' process design and implementation (including software development).
- Validation methods for integrity of preserved business processes.
- Overview of legal aspects for preservation of software services and their data.

5 Acknowledgements

Part of this work is funded by the European Commission's Seventh Framework Programme (FP7/2007-2013) under grant agreement No 269940.

6 References

[1] T. Kuny, “A Digital Dark Ages? Challenges in the Preservation of Electronic Information,” Sep. 1997.

[2] S. Strodl, P. Petrov, and A. Rauber, “Research on digital preservation within projects co-funded by the european union in the ict programme,” Vienna University of Technology, Tech. Rep., May 2011.

[3] CCSD, “Reference Model for an Open Archival Information System,” Consultative Committee for Space Data Systems, August 2009.

[3] D. Draws, S. Euteneuer, D. Simon, and F. Simon, “Short Term Preservation for Software Industry,” in *8th Int. Conf. on Preservation of Digital Objects* (iPRES 2011), Singapore, November 2011, pp. 1–9.

[4] D. Giaretta, “The CASPAR Approach to Digital Preservation,” The Internation Journal of Digital Curation, vol. 2, no. 1, 2007.