

# Processes for the Future: Sustainable External Dependencies of Processes

Stephan Strodl  
SBA Research  
Vienna, Austria  
sstrodl@sba-research.org

## ABSTRACT

Processes form an integral part of today's business and science world. They are realised by orchestration of services and tasks implemented by different information system. Nowadays, information systems are not any more self-contained, close system. They rather build on service oriented principles and integrate internal as well as external services and components to provide the required functionality and knowledge. These dependencies to third-party components pose a significant risk for the implemented processes. In particular, in the medium and long term the availability of external services cannot be guaranteed. Moreover, integrated third-party components can raise difficulties for modification and adaption for new requirements of the processes through restriction in licences or unavailable access to implementation sources. This thesis researches on the identification, assessment and preservation of third-party dependencies of processes. Dynamic and static approaches to identify dependencies of existing processes are analysed with a special focus on external dependencies. Based on a risk management approach the involved risks for the process are identified and assessed. Moreover, different non-traditional preservation approaches are investigated to preserve and maintain processes over time. Strategies for third parties dependencies are investigated. For example Software Escrow Agreements can be used to mitigate risks of software developed by a third party. The thesis describes the Escrow process and presents a framework to assess software development projects for deposit at an Escrow Agent.

## Categories and Subject Descriptors

D.2.4 [Software/Program Verification]: [Validation]; D.2.7 [Distribution, Maintenance, and Enhancement]: [Documentation]; D.2.9 [Management ]: [Life cycle]

## Keywords

Process Management, External Dependencies, Services, Digital Preservation

## 1. INTRODUCTION

Processes are increasingly supported by service oriented architectures, employing services offered by different, external providers. The use of external services allows to integrate required functionality and knowledge into information system to implement processes. The services are often offered as scalable and on-demand service models allowing organisation to optimize their resource consumption. These dependencies on third-party services pose a series of new challenges for the implemented processes. For the short term use of services the focus is on quality measurements such as performance, reliability and scalability, but other aspects become more relevant for the medium and long term such as availability and maintainability of external services and components. Software is in general not designed for long term availability, as it relies on a number of technologies for execution, for example hardware, operating systems, file formats and other software libraries, which all face the risks of obsolescence. External services are even more fragile as availability and functional stability cannot be guaranteed. Moreover the use of third-party software components can limit the adaptability and maintainability of the information systems. Adjustment of existing systems for new requirements or migration to other services can be excessively difficult due to missing access to source code or rights to modify used software components. This thesis addresses the challenges poses by process implemented by using third-party services and compartments. It consists of two parts, the first part investigates on the identification and assessment of third-party dependencies of processes. The second part of the thesis researches on preservation strategies to maintain processes for the medium and long run.

Businesses providing products and services with a long life-cycle have a need to maintain their data and processes over time to provide support for their products or to proof legal compliance in the future. Examples are business domains such as aviation, health, insurance and financial sector. Another driver for process preservation is science. Research intuitions and funding agencies want to ensure that the funded research is available in the future. Scientific institutions have also a vital interest in preserving their experiments, in order to verify the results in the future and to build on top of earlier experiments [8].

In order to keep the processes available, the supporting IT system and the required data need to be maintained over time or migrated to new systems. Information systems are

often implemented as service oriented architecture using external services and third-party components. Cloud computing for example has become a widely used technology offering different services that can be easily used and integrated into existing information systems. The use of cloud storage or external computation power increases the scalability and allows resource-efficient implementation of processes [4]. The financial stability of the provider, updates or replacements of the services can cause that services are no longer available in the future. But not only obvious dependencies through software services operated by externals involve risks for processes. Almost all software components are provided by external vendors, and subsequently the use of external software is regulated in terms of use, licences and contracts. Adoption and modification of the software for new requirements is often limited due to legal restrictions, refusal of the vendor or missing sources of the used software. The lack of ability to modify the software can hinder the maintenance of the processes over time and the implementation of new processes.

Both cases, the loss of external provided services and constraints of third-party software components, can have negative effects on the performance of the process. This thesis addresses the risks associated with the use of third-party services and software components for process implementations. The work has a particular focus on the medium and long term implications of external dependencies. The use of third-party services and software components leads to a dependency on the provider. This work investigates on the identification of external services for existing processes. Dynamic and static approaches are analysed with a focus on the dependencies between components and external services. A risk management approach is used to identify, structure and assess the risks associated with processes using different third-party services and components.

The thesis further investigates on preservation strategies for processes with a focus on approaches for third party services and components. The work considers well-established preservation strategies and novel approaches from outside the digital preservation domain to maintain processes and their external dependencies over time. Mock-ups are a promising candidate for external services and the concept of escrow agreements can be extended to deposit software service implementation from third parties. A framework is developed to support the assessment of deposit software projects for escrow.

The remainder of the paper is organized as follows. In Section 2 methods and tools to address the risk of using external services and components for processes are discussed. Section 3 looks at traditional and novel preservation strategies for process, with a special focus on strategies for third party services and components. A summary of the thesis's contributions is listed in Section 4.

## 2. THIRD-PARTY DEPENDENCIES OF PROCESSES

Process implementations often use and integrate services and components from third parties to provide the required functionality. Third-party dependencies can be distinguished into two delivery types: external services operated by third

parties and software components operated by the host organisation provided and developed by third parties. This thesis will address both types and investigate on techniques to identify external dependencies of existing processes and assess risk for processes in the medium and long term. It further investigates on dependencies between software components and their usage of external services (e.g. web services, external databases). The dependencies between components help to better understand the implementation of the process and is required to select appropriate risk measurements.

Different types of services are offered by third parties for customers that can be integrated and used to implement processes. Cloud computing technology has become popularity in the recent time [4]. Web services in general have become wide accepted for software development to integrate functionality. A set of Web service techniques like WSDL and SOAP<sup>1</sup> ease the use of external services in information systems. Besides Web services, databases services are offered for external customers. The knowledge base of expert systems is often realised as databases. It allows the integration of required external knowledge into expert system systems and to be used by processes. These systems have become very popular in medical and manufacturing sector and for business analysis [12]. Beside externally operated services, software components from external providers are used to implement processes. These components are operated by the host organisation but are licensed from third parties. Contracts and licences specify the allowed use and modification of the software. The operation of the service does not depend on the provider, but adaption, extensions and modifications can often only done by the developer of the software. Commercial and customised software is often delivered only as binaries by software providers. For modifications access to the source and the rights to edit it are required. Also the use open source licences has an impact on the future use of the software, e.g. in terms of selling and publishing. [9] discusses the legal aspects of software contracting and licences. A good example of used licenses can be seen in the list of used free and open source licences software used by Mercedes-Benz for the software for their cars [5].

In this thesis a set of real world processes from different domains is used to test and evaluate methods and tools for dependencies identification and assessment. These processes implement different third party dependencies, e.g. the use of external Web services, software from third parties and cloud-based application. The set includes amongst others a e-Science data process [14], a process for infrastructure construction models and a collaborative evaluation process. A clear understanding of the risks of external business relationships is inherent for organisations to ensure the availability and functionality of their processes over time.

### 2.1 Risk Management

The risk management approach in this thesis is based on the ISO 31000 standard [10]. The risk management helps to identify involved risks of processes in a structured way. It further supports the evaluation and selection of appropriate risk mitigation strategies. The effectiveness of applied

<sup>1</sup><http://www.w3.org/2002/ws>

strategies can be evaluated from a risk management perspective. Most related work has been done so far on short term risk of cloud computing with a strong focus on security issues [7]. Risks in digital preservation were focused on a data centric view [1, 2]. The TIMBUS project<sup>2</sup> addressed risk management approaches for business processes and e-Science processes [3, 15]. Based on the work of the TIMBUS project, we will investigate risks for processes with special focus on third-party relationships. Risks addressing availability, provided support, modifiability and functional stability of the external dependencies will be assessed. For the risk assessment, the significant properties [6] of the processes need to be specified. They help to identify the relevant behaviour and functionality that need be maintained over time.

## 2.2 Identification of third-party dependencies in existing processes

For the risk analysis all existing external dependencies of the process need to be known. A set of tools and methods will be investigated in this thesis to identify different external dependencies. The approaches aims at existing processes as process documentation is often outdated or incomplete. Static approaches can be used for example to analyse source code of software to identify external calls (such as Web services or databases) or identify third-party libraries [18]. Dynamic approaches allow to monitor process execution and identify the used resources of the process. Tools such Process Monitor<sup>3</sup>, Wireshark<sup>4</sup> and strace<sup>5</sup> will be evaluated by using the set of processes. The communication with external services needs to be identified. The type of service and the communication needs to be analysed and categorised. This thesis evaluates the capabilities of existing tools to identify and classify different types of dependencies. A framework will be developed that supports the identification and classification of external dependencies for processes.

To understand the dependencies of a process is essential for preservation. The framework is extended to detect dependencies between components and their connection to external services. Based on process monitoring the used resources and their external connections can be logged, analysed and represented in a model. The model allows an architectural view of the environment including the infrastructure, application and process aspects. The model builds the basis to select appropriate preservation strategies. The second part of the thesis as outlined in Section 3 investigates on new approaches for preservation of processes. A complete model of the process including resources, dependencies and external services is an essential prerequisite to maintain the process and its functionality over time. A key challenges is the capturing of the complete process rather than single process executions. It requires the execution and monitoring of all process execution paths to ensure that all components and connections are identified. The developed framework helps to monitor process execution, analysis the logs and create a model. The thesis evaluates on applicability, performance and shortcomings of this approaches based on a set

<sup>2</sup><http://timbusproject.net>

<sup>3</sup><http://www.sysinternals.com>

<sup>4</sup><http://www.wireshark.org>

<sup>5</sup><http://sourceforge.net/projects/strace>

of processes. Potential applications scenarios of this framework are wide-ranging and are outlined in the thesis, such as management of process evolution.

Besides the use of services operated by third-parties the thesis investigated on the use of third-party software that build the foundation of most information system. These software is regulated by licences and contract that specifies the allowed usage and potential modification of the software. New requirements can require adoption and modification of the process. The underlying software and its legal constrains need to be known for risk analysis and preservation. The thesis analysed exiting approaches identify the software and corresponding licences, e.g. Linux packages contain information about licences or the National Software Reference Library that contains haslets and metadata for software <sup>6</sup>.

The research questions for external dependencies are:

- What information is required to specify an external dependency with respect to its technical aspects?
- What are the risks for the medium and long term by using third-party services or software components?
- What static and dynamic techniques can be used to identify external dependencies in excising processes?
- What approaches can be used to identify external software components used in existing information systems?
- How far can existing process tracking tools can used to identify and classify external dependencies? What are the limitations of current tools?
- Which information can extracted from process monitoring logs to identify dependencies between components and their use of external services?
- What approaches exist to identify third-party software components used in process implementation?
- What approaches can be used to monitor external dependencies?

## 3. NOVEL PRESERVATION APPROACHES FOR PROCESSES

Work on digital preservation so far had a strong focus on static digital objects such as text and multimedia objects [17]. Recently, there is an increasing demand for preservation of processes from different domains. Business with long product lifecycles, such as aviation or health sector, have a need to maintain their processes and data over time. In order to provide service and maintenance for their products, old processes and data need to be available and executable. Processes represents an important asset in science, experiments are becoming more complex and more data intensive [8]. The verification and reuse of scientific experiments is a key motivation to preserve processes over time [11]. The shift in demands towards the preservation of complex processes and workflow can also be seen in the change of target content types of research projects in the area of digital preservation [17].

<sup>6</sup><http://www.nsrll.nist.gov>

Based on the TIMBUS Preservation Process Framework [16] we will investigate different preservation approaches to make processes available in the future or increase their robustness for future changes. A particular focus of this work will be on the preservation of third-party services and components. Traditional preservation strategies, such migration or emulation, will be evaluated, but we also consider approaches from outside the digital preservation domain to maintain processes over time, such as virtualisation or software escrow.

Different approaches will be investigated such as:

**Metadata/Documentation** helps to maintain the usability, interpretability, accessibility and understandability of processes. Particular manual steps of the process that are not executed by information system need to be documented for preservation.

**Migration** can be seen as the copying or conversion of digital objects from one technology to another. It is a widely adopted strategy for storage media and data formats. The migration of external services can help to solve third party dependencies. External services can be transferred into the own organisation (in-housing) ensuring the long term availability of the service. In-housing requires access to the implementation and data of the service as well as the rights to use the service. Example are cloud services such as storage or infrastructure that are operated by third parties.

**Emulation** mimics the behaviour and functionality of components, hardware or software. Emulation is a widely adopted strategy to preserve older computer platforms (e.g. video game console systems) and operating systems.

**Mock-up of components** provides a potential solution for third party software services (e.g. Web services) with no access to their implementation. A mock-up of the services is a form of simulation of the original service. The basic principle is to intercept and record messages from the original system to the service, which the simulation can then use to respond to request that have been captured previously. The approach is limited, as it can only be used for deterministic services (i.e. services for which the request and response pair always match, and which themselves are not dependent on any external state), and the mock-up can only respond to messages that have been recorded in the original system [13].

**Virtualisation** has become a common practice for management of IT systems. The most common version of hardware virtualisation provides a separation layer between the application services and the underlying hardware resources. The separation from actual hardware provides an abstraction layer of the physical environment. The use of virtualisation supports the portability of the systems. It also increase robustness against changes of the underlying hardware to a certain extent. Virtualisation often abstracts only part of the underlying hardware (such as storage), but uses other

resources like CPU directly. A combination of virtualisation and emulation can help to overcome hardware obsolesces, but using virtualisation to maintain software dependencies of complex information systems.

**SW-Escrow** offers a mitigation strategy to address third party service dependencies. It is often used for customised software. The basic principle is to place a third trustable party between the customer and the solution provider. Many organisations order development or customisation of software for their specific needs. The software is commonly provided as object code to the customer. For later adoption and adjustments of the software usually a service and maintenance contract is set up with the developer. This introduces a high dependency on the developer. In case of bankruptcy or refusal to maintain the program of the developer, the customer can have problems to further use the software, resulting in negative effects for his business. In a escrow case, all relevant artefacts of the software development are deposited at the escrow agent and released to the customer in case of predefined events. The deposit material needs to fulfill a certain degree of quality that allows the customer to further maintain and develop the software. To ensure this, all artifacts deposited are verified for their maintainability. Current business practice of software escrow does not consider verification of deposit material regarding its maintainability, bearing the risk that deposit material is incomplete and not useable for the customer in case of release. Deposit material for escrow often include only the source code that is insufficient to further develop and maintain the software. A software development projects consists of different artefacts, such as development environment, compiler, libraries, documentation, specification and data sources. [18] presents a framework to support the semi-automatic verification of software development projects for escrow deposit. The escrow approach can also be used for external services (such as Web services) provided by third parties. In this case the implementation of the service will be deposited at the escrow agent. If the operator of the service goes out of business, the user receives all required components to operate the service by himself. In this thesis we will evaluate verification measurements for software escrow, as presented in parts in [18], that can be used to assess the maintainability of software development projects.

The research questions for preservation approaches of processes and external dependencies are:

- What preservation strategies can be used to preserve processes?
- What information needs to be documented to reproduce a process instance?
- What strategies can be used to maintain external dependencies over time?

- What quality measurements for source code help to determine the maintainability of the software development projects for escrow?
- Which methods can be used to determine the completeness of software development artefact and how can third-party libraries be identified?
- What measurements can be used to assess the documentation of source code?

#### 4. CONTRIBUTION

The planned contributions of this thesis are:

- **Risk management approach for third-party dependencies of process**  
The thesis will provide guidelines to identify, assess and manage risk for third-party dependencies of processes.
- **Identification techniques for third-party dependencies of process**  
Techniques and tools are evaluated to identify the use of external services and software components of third parties.
- **Capturing and identification of dependencies between components and external dependencies of process**  
This thesis will investigate on the dependencies between components of process with a special focus on the interaction with external services.
- **Analysis of preservation approaches for processes**  
Different approaches are investigated to maintain the signification properties of processes over time with a special focus on approaches to address third party services and components.
- **Software Escrow Verification Framework**  
Identified shortcomings of current Software Escrow approaches will be addressed by developing an evaluation framework for software development projects. Evaluation criteria are specified to support assessing the maintainability of software development projects.

#### 5. REFERENCES

- [1] J. Barateiro, G. Antunes, M. Cabral, J. Borbinha, and R. Rodrigues. Digital preservation of scientific data. In *Research and Advanced Technology for Digital Libraries*, volume 5173 of *Lecture Notes in Computer Science*, pages 388–391. Springer Berlin Heidelberg, 2008.
- [2] J. Barateiro, G. Antunes, F. Freitas, and J. Borbinha. Designing digital preservation solutions: a risk management based approach. *The International Journal of Digital Curation*, 5:4–17, 2010.
- [3] S. Canteiro and J. Barateiro. Risk assessment in digital preservation of e-science data and processes. In *Proceedings of the 8th International Conference on Preservation of Digital Objects (iPRES 2011)*, 2011.
- [4] N. Carr. *The Big Switch: Rewiring the World, from Edison to Google: Our New Digital Destiny*. Norton, W. W. & Company, Inc., 2008.
- [5] Daimler AG. Licence agreement supplement. Technical report, [http://www4.mercedes-benz.com/manual-cars/ba/foss/content/en/assets/FOSS\\_licences.pdf](http://www4.mercedes-benz.com/manual-cars/ba/foss/content/en/assets/FOSS_licences.pdf), 2012.
- [6] A. Dappert. Deal with conflict, capture the relationship: The case of digital object properties. In *Proceedings of the 7th International Conference on Preservation of Digital Objects (IPRES2010)*, pages 21–29, September 2010.
- [7] B. Grobauer, T. Walloschek, and E. Stocker. Understanding cloud computing vulnerabilities. *Security Privacy, IEEE*, 9(2):50–57, 2011.
- [8] T. Hey, S. Tansley, and K. Tolle, editors. *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Microsoft Research, Redmond, Washington, 2009.
- [9] T. Hoeren, B. Kolany-Raiser, S. Yankova, M. Hecheltjen, and K. Hobel. *Legal Aspects of Digital Preservation*. Edward Elgar Publishing Ltd, 2013.
- [10] ISO. *ISO 31000: 2009 Risk management – Principles and Guidelines*.
- [11] D. Koop, E. Santos, P. Mates, H. T. Vo, P. Bonnet, B. Bauer, B. Surer, M. Troyer, D. N. Williams, J. E. Tohline, J. Freire, and C. T. Silva. A provenance-based infrastructure to support the life cycle of executable papers. *Procedia Computer Science*, 4:648 – 657, 2011.
- [12] C. T. Leondes, editor. *Knowledge-Based Systems, Four-Volume Set: Techniques and Applications*. Academic Press, 2000.
- [13] T. Miksa, R. Mayer, and A. Rauber. Ensuring sustainability of web services dependent processes. *International Journal of Computational Science and Engineering (IJCSSE)*, 2013. Accepted for publication.
- [14] M. A. Neumann, H. Miri, J. Thomson, G. Antunes, R. Mayer, and M. Beigl. Towards a decision support architecture for digital preservation of business processes. In *Proceedings of the 9th Int. Conf. on Digital Preservation (iPres 2012)*, 2012.
- [15] D. Simon, J. Barateiro, and D. Burda. Leveraging DP in commercial contexts through ERM. In *Proceedings of the 10th International Conference on Preservation of Digital Objects (IPRES 2013)*, pages 104–109, Lisbon, Portugal, 2013.
- [16] S. Strodl, R. Mayer, G. Antunes, D. Draws, and A. Rauber. Digital preservation of a process and its application to e-science experiments. In *Proceedings of the 10th International Conference on Preservation of Digital Objects (IPRES2013)*, pages 128–136, Lisbon, Portugal, September 2013.
- [17] S. Strodl, P. Petrov, and A. Rauber. Research on digital preservation within projects co-funded by the European Union in the ICT programme. Technical report, Vienna University of Technology, May 2011.
- [18] E. Weigl, J. Binder, S. Strodl, B. Kolany, D. Draws, and A. Rauber. A framework for automated verification in software escrow. In *Proceedings of the 10th International Conference on Preservation of Digital Objects (IPRES 2013)*, pages 95–103, Lisbon, Portugal, September 2013.