SERIES	TIMBUS WHITE PAPERS
Title	THE TIMBUS PROCESS

TIMBUS WHITE PAPER: THE TIMBUS PROCESS

Publication Date: 28 November 2014

Dissemination Level: PU

TIMBUS is supported by the European Union
under the 7th Framework Programme
for research and technological development and demonstration activities (FP7/2007-2013)
under grant agreement no. 269940

SERIES	TIMBUS WHITE PAPERS
Title	THE TIMBUS PROCESS

SERIES	TIMBUS WHITE PAPERS
Title	THE TIMBUS PROCESS

Authors		
Name	**Organisation**	**e-mail**
Stephan Strodl	SBA	sstrodl@sba-research.org

Contributors		
Name	**Organisation**	**e-mail**

Internal Reviewer		
Name	**Organisation**	**e-mail**
Artur Caetano	INESC-ID	artur.caetano@tecnico.ulisboa.pt

Disclaimer

Table of Contents

List of Figures

1 Executive Summary

Digital Preservation has predominantly focussed on static objects, such as data sets, documents and media. However, there is an increasing demand on preserving the processes that manage and use those data sets for the long term, such as scientific workflows and business processes. The TIMBUS research project aims at providing a methodology, guidelines and tools to digitally preserve business processes. The TIMBUS process framework defines the preservation process and actions needed for planning the preservation, executing the preservation, and redeploying a process in a new environment at some point in the future. The TIMBUS process framework is domain independent and can be adjusted and applied to different usage scenarios. The framework describes the information flow between the steps of the TIMBUS preservation process, provides recommendations regarding suitable methods and tools needed for performing the preservation, and defines responsibilities and roles of stakeholders involved in the different preservation steps. The TIMBUS process consists of three phases: plan, preserve and redeploy. TIMBUS adopts a risk management approach that considers Digital Preservation as a risk mitigation strategy on an enterprise wide perspective for business processes. TIMBUS contributes to Digital Preservation by adding long-term availability and usability qualities to business processes, along with its data, supporting services and infrastructure, legal and organisational aspects.

2 Introduction

Digital Preservation is the management of digital information over time. While the research on Digital Preservation traditionally follows a data-centric view of information, the long-term preservation of processes and supporting infrastructure has received less attention. Currently there is an increasing interest in the preservation of processes, going beyond the traditional data preservation. The TIMBUS project aims at providing a methodology, guidelines and tools to capture, document, and preserve processes for the long term. This white paper presents the three process steps of the TIMBUS process framework, namely plan, preserve, and redeploy.

A business process is a set activities orchestrated to achieve a specific and well-defined goal. The preservation of a business process requires sufficient details of the process structure as well as the context it is embedded in. The context situates the process within its system of use and captures its dependencies, stakeholders and qualities, such as restrictions and requirements. While the process structure is sufficient to understand and execute its workflow, the context is needed to redeploy the process with its original behaviour as exhibited in the original system of use. Moreover, detailed planning and testing of potential preservation strategies, followed by a controlled execution of the required preservation actions, is needed for a correct future redeployment of the process. Process preservation needs to address the different layers that a process crosscuts, including the business, application, and technology layer. The relevant context of the process has to be captured including technologies both on hardware (technology) and software (application) levels, dependencies between these elements, use of external services, data, and high-level concepts pertaining to the business layer. The business layer addresses the organisational and operational aspects of a business process including its business logic, the involved stakeholders and legal obligations. Business processes may be totally or partially automated by services provided by different information systems, within or outside the boundaries of the system owning the business process. Since such external services are beyond the organizational boundary, their capture and preservation is not straightforward as they are outside the controllable domain. In the long run, the availability of current technology, both hardware and software, as well as services cannot be guaranteed, especially if they are provided by third parties. Thus, the inability to preserve software, services and technology threatens the preservation of a business process and its future redeployment. Legal concerns, such as non-compliance with contracts, licences, patents or SLAs (Service Level Agreements), constitute a further threat to process preservation and redeployment. Moreover, domain specific laws and regulations such as data protection set further requirements and obligation for a preservation solution.

The TIMBUS process framework guides organisations through the three phases of the preservation approach. The approach is driven from an enterprise risk management perspective. In this perspective, Digital Preservation is a risk mitigation strategy that targets the threat of loss of availability of information over time. Here, risk management is used to identify and evaluate and structure risks associated with business processes preservation and redeployment. This approach

enables different preservation alternatives to be properly identified and evaluated, thereby contributing to making an informed decision regarding the preservation strategy.

Within the planning phase, the context of the process is captured, including all relevant aspects that are needed for its preservation and future redeployment. This phase produces a preservation plan with the relevant actions that need to be executed for extracting the relevant process data from the productive environment, and subsequently preserving and archiving the process for the long term. The relevant components implementing and documenting the significant properties of the process need to be identified and maintained over time. Processes may not be limited to a single system but often make use of distributed services across platforms that need to be properly identified. The used components need to be described and preserved for future use. Based on cost and risk considerations, suitable preservation strategies are then selected. The preservation phase executes the preservation strategies and prepares the process for long term archiving ready to be redeployed. The redeployment of the business process defines the re-execution of the preserved process in a new environment at some point in the future with the goal of duplicating the original context of operation. The redeployed process needs to be verified for its authenticity and correctness against the original process and context. An evaluation is required to compare the significant properties of the redeployed process against reference values from the original process.

The TIMBUS process describes the required actions that need to be performed and the expected outcome. Responsibilities at process level are defined and its stakeholders are specified. The TIMBUS framework is not limited to a single domain-specific use-case but provides the flexibility to be applied in various settings. Specific tool support for the aforementioned actions is part of the TIMBUS project and is available at https://opensourceprojects.eu.

Details about the TIMBUS process framework can be found in [1] and its application is shown in [2].

3 TIMBUS Digital Preservation Process Framework

The TIMBUS Digital Preservation Framework defines the process steps to digitally preserve a business process according to three phases: planning, preservation and redeployment. The process steps of each phase specify its inputs, outputs, methods and responsibilities. The goal of this framework is to guide organisations through the process of preserving business processes. The detailed description of the framework is presented in [1]. The framework is domain independent and can be applied to different settings by adjusting to the needs and requirements of the specific scenario. The instantiation of the framework needs to be done for specific settings depending on the implementation of the process, requirements, obligations and involved actors. It provides a clear specification of the actions and information that is required for preservation and redeployment of a business process, but is flexible enough to be adjusted according to the requirements of specific settings. Figure 1 shows the high level processes of the TIMBUS framework and their dependencies on risk management. The TIMBUS process can be divided into three phases: plan, preserve and

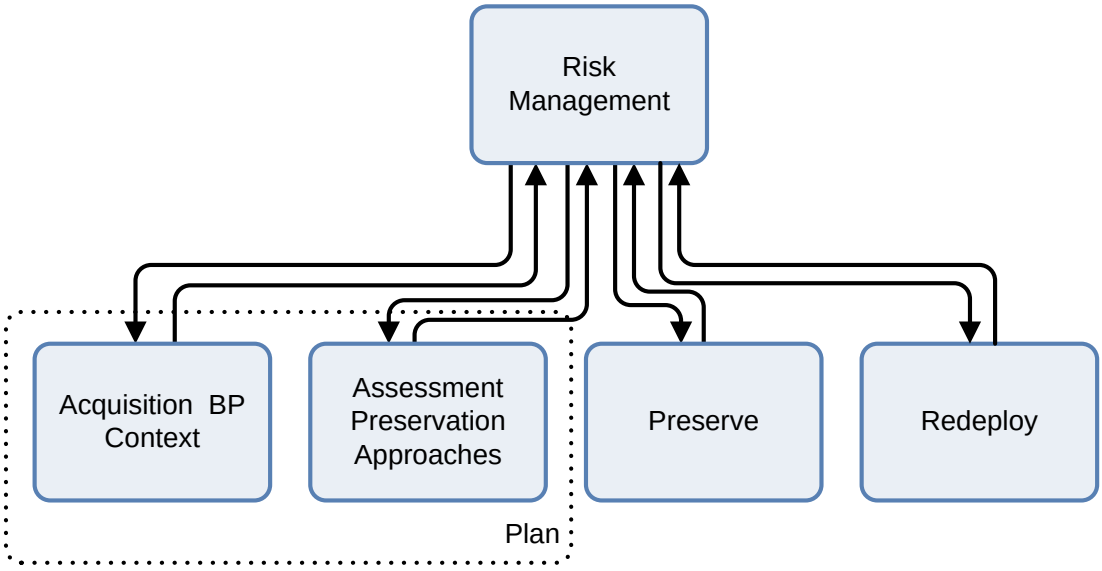

Figure 1: High level view of the TIMBUS Framework process

redeploy.

In the first phase, plan, risk management provides a risk analysis of existing business processes including an assessment of the processes that need to be accessible in the future. In order to assess the risks of a process, its context and dependencies are captured and modelled (this action is named *Acquisition of the business process context*). Processes that display a high risk of becoming unavailable

require mitigation strategies that minimize the preservation risk. This is accomplished by the *Assessment of Preservation Approaches*. This includes processes with external components and those which are beyond the organizational boundary.

The actual execution of the preservation strategies and transformation of the business process into an archival format is done by the *Preserve* process. The *Redeployment* process enables the archived business process to be executed in a new environment at some time in the future.

Figure 2 expands Figure 1 and shows a detailed view of the TIMBUS process outlining its major sub processes and its three main actors: Risk Manager, Preservation Manager and Preservation operator. . The next sections describe each of the three process phases.

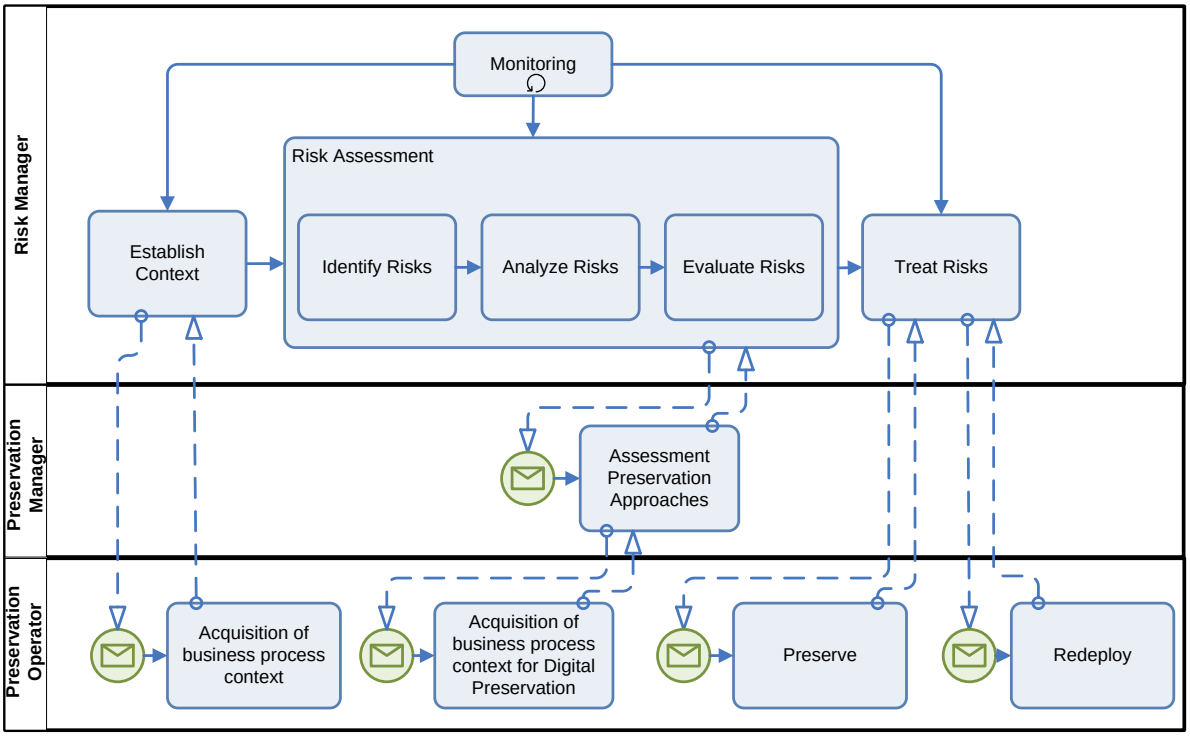

Figure 2: Detailed view of the TIMBUS Procsess framework

3.1 Planning phase

The planning phase is responsible to capture the process and its context as well as to assess suitable preservation approaches. As shown in Figure 2, the first step is the acquisition of the process context, followed by risk assessment. Risk assessment triggers the assessment of preservation approaches required for the identification, specification and evaluation of preservation strategies.

3.1.1 Acquisition of the business process context

To successfully capture the context of a business process, the framework makes use of a context model that enable to systematically capture the essential aspects of the process required for its preservation, verification and later re-execution. The context model is described in detail in [3] [4] [5] [6] [7]. The model is realized as an ontology, described with the OWL language, which facilitates performing reasoning, analysis and conformance checking.

As the context of a process usually involve a variety of heterogeneous concepts, such as software, hardware, people and legal obligations, the meta-model that underpins the context model must be able to cope with not only with generic concepts that apply to several domains but also with domain-specific aspects. This challenge is addressed by using an upper ontology that deals with the domain-independent concepts, named domain-independent ontology (DIO) that provides the generic concepts, and a set of domain-specific ontologies (DSOs) that are integrated and mapped to the DIO. The DIO is grounded on existing work in enterprise architecture as it provides the constructs to handle the core concepts and dependencies required to describe business processes and their dependencies and realizations at multiple levels of details and from multiple viewpoints, such as business, applications and infrastructure. Specifically, we adopted ArchiMate [8], which specifies a language and a framework that describes an enterprise architecture using a set of five core concepts that are specialized with a total of around 30 different concepts describing business, application and technological aspects.

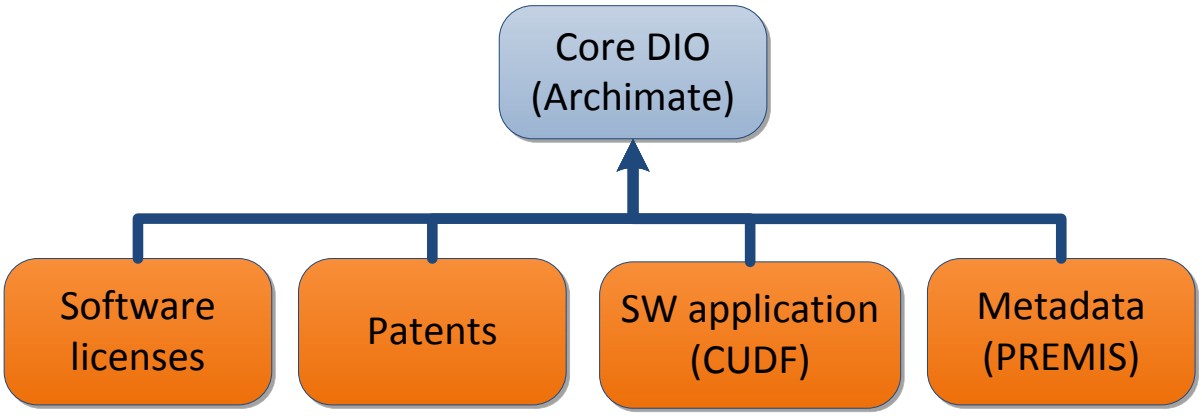

Figure 3 Example of a context model with the ArchiMate DIO and four domain-specific DSO.

Figure 3 shows the mapping structure of the context model. We developed a number of domain specific ontologies (DSO) including:

- Software licenses, based on The Software Ontology[1],

[1] http://theswo.sourceforge.net

- Patents, based on the Patent Metadata Ontology (PMO), developed by the PATEXpert project[2],
- Software application dependencies, based on the CUDF, the Common Upgradeability Description Format [9],
- Digital Preservation meta-data, based on the PREMIS data dictionary [10].

Some elements of these domain-specific ontologies are identified as sub-types of concepts defined in the DIO, and are mapped to these respective elements. This allows for a comprehensive description of the domain-specific aspects, while keeping the core ontology minimal and proving traceability between the concepts.

However, capturing the context of a specific use case may require using concepts that are not part of the DIO and the DSOs described above. To do so, the context model can be further extended with other DSOs that define domain specific aspects of the use case. Some parts of these DSOs can be acquired automatically, such as the software dependencies on package-based operating systems such as Debian Linux, which also provides means to identify the licenses a certain package is distributed under. Other elements will have to be provided manually, for which we provide a graphical editor, implemented as a plugin to the Protégé ontology editor[3] and which is part of the TIMBUS toolset.

3.1.2 Risk Management

Risk management has the goal of defining prevention and control mechanisms to address and mitigate risks related to assets and activities. Preservation can be seen as a potential method to mitigate risks, derived from the potential loss of information over time. The risk management process used in TIMBUS is based on the ISO 31000 standard [11]. TIMBUS defines the process-related interfaces to connect Digital Preservation with risk management. From the TIMBUS perspective, the risk associated to a process can act as a driver prompting its preservation as a way of mitigating the threats that endanger it. Thus, risk management helps to identify and evaluate different risks in a structured and well defined manner. If risks related to information have been identified and evaluated, different preservation alternatives need to be developed. The risk management process facilitates making a decision about the mitigation strategy. Whenever process preservation is considered to be the suitable risk treatment, the preservation process is triggered. The TIMBUS risk management approach is described in more detail in [12] [13].

Assessment of Preservation Approaches

[2] http://www.patexpert.org

[3] http://protege.stanford.edu

The Assessment of Preservation Approaches process is responsible for the identification and evaluation of different preservation approaches. It starts with the refinement of the context model as shown in Figure 2. In the first iteration the context model was created for the use by risk management related activities. However, more information about the technical implementation of the process is required for the planning of preservation strategies.

The preservation requirements of the process are specified and documented for the evaluation and comparison of preservation approaches. The requirements specify the significant properties, describing functional and non-functional requirements of the process that need to be maintained over time. Redeployment scenarios support the specification of the significant properties regarding the preservation of artefact and execution of the process in terms of performance and behaviour. Different redeployment scenarios for future usage can be considered, e.g. execution of the original process with original data for confirmation of documented outcomes, execution of the original process with new data, or to modify parts of the process but using the original data e.g. for scientific workflows to evaluate improvement of new methods or models on the experiment results.

Other preservation requirements can include amongst other, checking the compliance against standards, institutional policies or legal obligations. The requirements are also later used to evaluate and compare different preservation approaches as well as to verify the redeployment at a later stage. More details about the specification of signification properties can be found in [14].

A process is an orchestration of tasks that are executed in a particular sequence. Its execution can involve using different services from various systems. For this reason, a combination of different preservation actions can be applied to preserve a process for the long term. An examples is using virtualisation to preserve the functionality of a set of services and its software and hardware infrastructure, emulation to preserve the functionality of services while disregarding the underlying infrastructure, and data migration to support document preservation.

A challenging task for the preservation of complex processes is the ability to preserve relationships and dependencies between components over time. Knowledge of the dependencies is important for maintaining the functionality of the components. Broken dependencies can prevent the redeployment of the process in the future. Examples are manifold, such as missing libraries for software execution, missing databases for data input, incompatible hardware for operating systems or missing credentials for encrypted data. The dependencies need to be considered whenever changes are applied to components. Modification of components for preservation purposes for example can have undesired side effects on other components. Examples are the migration of data into other formats that cannot be processed further by other software components, or the replacement of software components by new versions that offer different interfaces for interaction. Reasoning and queries based on the context model can help to identify dependencies and to further support finding feasible preservation approaches.

While strategies for Digital Preservation, so far, mainly focus on data migration and emulation, the preservation of processes need further approaches, especially with respect to external dependencies. Different strategies can be used to maintain the significant properties of the process over time. Examples of strategies that support preservation and archival of processes are:

Metadata/Documentation

In order to maintain the usability, interpretability, accessibility and understandability of the process, additional metadata of its components is required. Understandability involves providing sufficient information so the component can be interpreted and understood in the future. Manual steps of the process that are not implemented by information systems require documentation for later redeployment. Furthermore logging and tracking functionalities of software components (such as workflow engines) can be used to document the process execution and provide provenance information for the future.

Migration

Migration can be seen as the replication or conversion of digital objects from one technology to another. It is a widely adopted strategy for storage media and data formats. Besides that, the migration to alternative software services or components can be applied to process preservation. For example in terms of software licences, the use of alternative open source resources that provide the same functionality can be a suitable strategy to overcome legal conflicts. Another aspect of migration relates to the use of services provided by third parties. As the availability of external services cannot be guaranteed, a potential strategy is to transfer these external services into the process' own system (in-housing). The strategy requires access to the implementation and data of the service as well as the licences and rights to operate the service. An example is cloud storage services that are operated by third parties.

Emulation

An emulator software 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.

Virtualisation

Virtualisation, especially hardware virtualisation, has become a common practice for infra-structure and server management. Virtualisation software provides a separation layer between the application services and the underlying hardware resources. The separation from actual hardware abstracts the physical environment, such as network, storage and display. As so, it increases the robustness of virtual machines (VM) against changes of the underlying hardware, with the cost of adding an

additional abstraction layer. The virtualisation is a practical approach to capture complex systems while maintaining their dependencies fully enclosed within a VM.

Mock-up of Software Services

A problem for preservation is the use of third party software services (e.g. web services) within a process. A potential solution is a mock-up of the services in form of a simulation of the original service. The basic principle is to intercept and record messages from the original system between process and service, which the simulation can then use to respond to request that have been captured previously. As such, it is a restricted form of emulation in the sense can only respond to messages that have been recorded in the original system. Moreover, mock-ups can only be used to replace deterministic services, i.e. services for which the request and response pair always match, and which themselves are not dependent on any external state. Nevertheless, for deterministic services and for the preservation of particular instances of a process, the mock up can actually provide a suitable and cost-effective solution despite its limitations. An analysis of mock-up strategies for web services, and recommendations to make Web services more resilient in general, can be found [15].

Software Escrow

Processes are often implemented using proprietary and customised software applications and services. The software is in many cases delivered as closed source to the customer, meaning the source code remains at the vendor and only the binaries of the software are delivered to the customer. From the preservation perspective, this scenario limits the potential preservation strategies for the software, as the software cannot be adapted to future changes in the execution environment. Software Escrow offers a mitigation strategy as it places a trustable third party between the developer and the customer [16]. All artefacts relevant to the software development are deposited at the escrow agent and released to the customer in case of predefined events, e.g. when the vendor goes out of business, or does not want to further maintain the software.

Different approaches can be used to preserve a process, using different strategies or tools. Each approach is specified in a *Process Preservation Plan*. The plan also defines procedures for capturing the process data and later redeploying and verifying the process. In order to preserve the process, the components and process data need to be captured from the source systems. The acquired data needs to be in a consistent state so that redeployment leads to a valid state of the process, e.g. all database transaction are closed. The redeployment procedure defines the execution of the preserved process in a new environment. In order to ensure that the process is redeployed correctly, a verification and validation procedure is required. This procedure defines measurement points to check that the redeployed process displays the same significant properties as the original process.

The proposed plans are evaluated against the previous specified preservation requirements. The evaluation includes assessing whether the proposed models and procedures are complete and correct

and that all significant properties of the process are preserved. In case the evaluation shows that relevant aspects of the process are missing or requirements are not fulfilled, a feedback loop of the *Assessment of Preservation Approaches* process allows the refinement of the context model or the preservation plan specification until the model complies with the requirements. With this approach, different preservation plans can be evaluated, and the evaluation results are submitted to the risk management for decision making. The impact of the different strategies on identified risks is assessed and the best matching solution is selected for treatment.

3.2 Preservation phase

The acquisition and preservation procedures of the *Process Preservation Plan* are applied to the business process in the preservation phase. The software and data of the process are captured from the source environment. Preservation actions are executed and the process is prepared for archival storage. Validation and verification data are captured from the source system for redeployment [14].

3.3 Redeployment phase

The redeployment phase defines the reactivation of a previously preserved process in a new environment at some point in time. The key characteristics of the new environment are captured, including the available technical components, organisational and legal aspects. A gap analysis between the requirements of the preserved process for redeployment and the available environment is then performed. The technical infrastructure needs to be adjusted and prepared for the redeployment, including different approaches to overcome identified gaps, e.g. emulating components or services, or migrating data. Required software and data are installed according to the redeployment procedure defined in the *Process Preservation Plan*. Tools and components for validation and verification are also set up in the new environment. As a final step, the process is re-executed and the measurements obtained from its execution become available for validation [14].

4 References

[1] TIMBUS consortium, “D4.6: Use Case Specific DP & Holistic Escrow,” 2013.

[2] 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)*, Lisbon, Protugal, 2013.

[3] TIMBUS consortium, “TIMBUS Whitepaper: Context Model,” 2014.

[4] TIMBUS consortium, *D4.2: "Dependency Models Iter. 1"*, 2012.

[5] TIMBUS consortium, *D4.3: "Dependency Models Iter. 2"*, 2013.

[6] R. Mayer, T. Miksa and A. Rauber, “Ontologies for describing the context of scientific experiment processes,” in *10th International Conference on e-Science*, 2014.

[7] TIMBUS consortium, *D4.5: Business Process Contexts,* 2012.

[8] The Open Group, “ArchiMate 2.0 Specification,” 2012.

[9] R. Treinen and S. Zacchiroli, “Description of the CUDF,” 2008.

[10] PREMIS Editorial Committee, “Premis data dictionary for preservation metadata,” 2008.

[11] ISO, “ISO 31000: 2009 Risk management - Principles and Guidelines”.

[12] TIMBUS consortium, “TIMBUS Whitepaer: Risk Management,” 2014.

[13] TIMBUS consortium, “D4.8 Refined DP & Intelligent Enterprise,” 2013.

[14] TIMBUS consortium, “TIMBUS Whitepaper: Verification and Validation of Preserved and Redeployed Business Processes”.

[15] T. Miksa, R. Mayer and A. Rauber, “Ensuring sustainability of web services dependent processes,” *International Journal of Computational Science and Engineering (IJCSE)*, 2013.

[16] 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)*, Lisbon, Portugal, 2013.