Applying the TIMBUS Approach to Preserving Context in Digital Libraries

Carlos Coutinho
Caixa Magica Software Rua Soeiro Pereira Gomes, Lote 1-4°B Edifício Espanha, 1600-196 Lisboa +351 217 921 260
carlos.coutinho@caixamagica.pt

Paul Gooding
Digital Preservation Coalition
British Library, Floor 5, Room 14
96 Euston Road, London, NW1 2BD +44 (0) 20 7412 7329
paul@dpconline.org

ABSTRACT

To date, digital preservation has generally focused on the preservation of specific data in the form of artefacts. However, in order to ensure that this data is accessible in the long term, it is vital that we consider how to extract and preserve information on the software and hardware contexts which this data depends upon to operate. We therefore need tools to assist in identifying, preserving and validating the processes which underpin the creation of data in digital libraries.

In particular, we need to consider the importance of preserving not just individual digital artefacts, but the platforms which allow digital libraries to render or execute their items. Digital libraries rely on this software to render their items, and it is therefore important to know configuration details and software dependencies to ensure these items remain fully operational in the future. In the case of digital libraries, the TIMBUS framework provides the tools necessary to assist practitioners in identifying relevant processes, undertake risk analysis, and then to assist the user in extracting, preserving and revalidating the necessary processes.

This half-day tutorial introduces the TIMBUS approach to process preservation, and demonstrates how it can be applied to issues relating to digital libraries. TIMBUS focuses primarily on business processes, but this tutorial will show its approach to process-oriented preservation is also relevant to digital libraries. It provides a methodology for process preservation and a set of tools which help to semi-automatically validate and preserve processes so that they can be recreated at a later dat. Participants will be given the knowledge to understand the importance of technical environments for collection items, and learn more about the TIMBUS solutions through examples relevant to the digital library domain. They will also gain an understanding of digital preservation as a risk mitigation strategy.

1. TUTORIAL LEVEL
Introductory level.

2. DURATION
Half-day (3 hours).

3. PARTICIPANT NUMBERS
Up to 20.General Terms

4. OUTLINE OF THE CONTENT
The tutorial will cover the following topics:

- An introduction to the basics of digital preservation: here the purpose is to present an overview of the state-of-the-art in digital preservation techniques and motivations, and to equip participants with the requisite knowledge for the tutorial;

- Discussion of moving from data-oriented preservation to process-oriented preservation: this section aims to show the problems regarding current approaches to business preservation. Whereas up to now the main concern was to capture the business data (e.g., databases and logs), the paradigm is shifting to understanding the underlying business processes and, besides modelling and capturing, to preserve the complete surrounding environment;

- Explanation of the TIMBUS approach to process and context preservation by presenting the TIMBUS storyboard;

- Introduction to the TIMBUS set of tools and showcase of the TIMBUS architecture model, including:
  - Context Capturing tools;
  - Context Model tools;
  - Risk Management tools;
  - Digital Preservation Expert Suite, Preservation Identifier and Dependencies Reasoner;
  - The TIMBUS methodology for process preservation;

- Presentation of a Context Model for capturing and describing processes: this will define the context model ontologies (Domain Independent Ontologies and Domain-Specific Ontologies) and corresponding support tools, including Archi, Archi2OWL, Jena, Protégé;

- Discussion of the challenges of automatic and semi-automatic capture of context, including definition of the context capturing tools;

- Explanation of how to adapt tools and models to the heterogeneity of systems and businesses, including a showcase of the paradigm of development by

contributions and architecture off the context metadata capturing tools using OSGi;

- Explanation of process capture and modelling from real business artefacts, including definition of tools used to capture and model business processes based on real business evidence sources such as logs;

- Introduction to variants of risk analysis: Classical and Simulation. The suite of risk management tools in TIMBUS will be introduced, including their features and approaches towards:
    - o Classical Risk Management: Risk Evaluation and Treatment tool;
    - o Simulation Risk Management: Intelligent Enterprise Risk Management tool (iERM);

- Explanation of Digital Preservation as a risk management strategy. This will present the Digital Preservation Expert Suite of tools for performing Digital Preservation as one solution provider for Risk Management mitigation.

5. INTENDED AUDIENCE

The tutorial is aimed at researchers, publishers and curators in digital libraries, who want to learn about process-oriented digital preservation. Some understanding of digital preservation will be helpful, but the tutorial will begin with an introduction to the basics.

6. EXPECTED LEARNING OUTCOMES

The tutorial participants will gain understanding of the concept and importance of process-oriented preservation. They will learn how to adapt tools and models for this purpose, how to capture processes, and the basics of risk analysis for process preservation. They will learn more about the TIMBUS suite of tools, and be introduced to the role of these tools in successfully preserving the entire business environment, including both data and business-oriented preservation. As a result, participants will be empowered to explore these issues in relation to their own organisations.

7. BIOGRAPHY OF THE PRESENTER(S)

Carlos Coutinho is a Senior Research engineer and R&D Project Manager at Caixa Mágica Software in Lisbon Portugal. He holds a PhD degree in Electrical and Computer Engineering (2013), awarded by the New University of Lisbon (FCT-UNL), Portugal, where he also does research, with interests in Enterprise Interoperability, Adaptable Platforms and Systems, SOA, and Model-Driven Engineering. He has more than ten years of experience teaching the fields of IT at Portuguese universities ISEL, ISCAL, ISGB and ISCTE. He has five publications in international scientific journals (with ISI-IF) and more than fifteen publications in peer-reviewed international conferences, and is part of the review committee of two journals and three yearly conferences. He also holds a PMI-PMP© title and has a post-graduation in Project Management by Instituto Superior Bissaya-Barreto (ISBB) in Coimbra, Portugal. He has more than fifteen years of experience working as an engineer in the enterprise IT area, working in several fields from ICT, Services, Public Administration and the Aerospace industry, in several multinational companies like Alcatel, Siemens and Critical Software.

Other presenters to be confirmed.

8. Workshop Outcomes

- A very important outcome that was intended with this workshop was for the TIMBUS development team to have a clear feel of the impact of the project's tools on the final customers and on the potential users that are going to support the project outcomes. More than producing outputs that fill the objectives of the project on the EC reviews, TIMBUS intends to produce tools that are in fact usable for DP customers and practitioners;

- The community that assisted the workshop was indeed quite interested in the outcoming tools that were produced and showcased, particularly on the context model ontologies and on the business process extraction framework, as well as on the whole concept of risk-driven digital preservation.