Digital Preservation of Business Assets as a Risk Management Strategy

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

ABSTRACT

The effect of a new globalised world where technologies and methodologies mature rapidly is a constant change in business paradigms and assets. Tools, technology and formats become easily deprecated and replaced by others, and legal and licensing affairs are subject to constant legislation changes. Hence, a business seeking to preserve its assets for long term becomes very prone to problems of this nature. This paper presents an approach to analyse the risks coming from various sources which may affect enterprise business continuity, discussing the problems and issues relevant to it, and proposing long-term digital preservation as a mitigation action to handle some of these determined risks, and the implementation of a framework to support these concepts, in the context of the Seventh Framework Programme (FP7) TIMBUS project.

General Terms

Management, Documentation, Design, Experimentation.

Keywords

Risk Management, Digital Preservation.

1. INTRODUCTION

In a globalised and networked world, businesses face the struggle to improve, to perform better and faster, using new approaches and techniques and optimising their way in order to stay in business. This means that processes and environments keep changing, rapidly and deeply.

To be able to acquire a business culture, as well as to be able to replicate past behaviours, this need to change must be coped with an equally thorough concern about making sure all the produced business artefacts and flows are able to be persisted and perceived in long term. The methodologies and tools developed in the EU funded research project TIMBUS [9] constitute a key element to enable these processes.

OSDOC'13, July 11-12, 2013, Lisbon, Portugal.

Enterprises have as core their business assets. These assets concern various forms and formats, including e.g., abstract concepts, flows, strategies, hierarchies, culture and tacit knowledge, to more down-to-earth files, records and other business evidences. The path to success on enterprises includes storing as much and as detailed information as possible, in order to be able to infer conclusions, tendencies, behaviours and needs from this information. In a recent interview, the IT director from Virgin Atlantic stated their new 787 fleet is expected to create more than half terabyte of data for each flight [4]. As businesses grow every day, so does the amount of information that is produced. In a study conducted by IDC [5], digital archive growth will reach data amounts of 40 ZB (Zettabytes) by 2020.

Hence, being able to ensure business continuity involves several steps and stages, starting on a major effort on serialising and modelling the abstract concepts and knowledge into storable information and then ensuring all the stored information is maintained. Maintaining knowledge is a challenge on its own: not only is the stored information prone to physical corruption of the storing media, but also this media needs to be adequately maintained so that means can be found to reinstate the knowledge.

Formats and versions keep updating, businesses get discontinued, and finally, there is the need to cope with the underlying change that is required to evolve. The adoption of new standards in information management and modelling, techniques, trends and formats (e.g., Helium storage disks [7], genetic hard drives [2], files formats), information transmission and connectivity.

This together with the new paradigms on semantic analysis and knowledge capturing, it all contributes to harden the maintenance that is needed to provide business continuity, making it a real challenge to distinctively meet the objective to provide medium to long term preservation and disaster recovery while preserving and storing business processes and supporting software and technology stack.

TIMBUS stands for “Digital Preservation for Timeless Business Processes and Services”, and focuses on having resilient business processes. In order to maintain the business continuity of an enterprise, TIMBUS has the objective to ensure continued access to services and business assets, providing processes and methods for digitally preserving business processes. It includes a flexible and extensible software architecture, intelligent tools and technologies for performing the major functionalities of business and knowledge capturing, analysis and maintenance, preservation and storage of the information and meta-information, and future redeployment of the stored information to knowledge.

Section 2 provides the proposed solutions for dealing with the above described problem and the status of the implementation of these solutions, section 3 will present the project's validation use-cases, and section 4 will show the retrieved conclusions and future work to be taken.

2. PROPOSED SOLUTIONS

Digital Preservation, as an academic discipline and organisational practice aims to ensure the availability of information over a long period of time. It is essentially motivated by business and legal risks incurred by information loss or damage. On the other hand, Risk Management is traditionally addressed as a management discipline, performed typically in an isolated fashion within organisations, lacking the perspective on business processes. The proposed solution undertaken by the TIMBUS project extends the traditional digital preservation approaches as it introduces the need to analyse and sustain accessibility to business processes and the supporting services, aligning preservation actions with enterprise risk management (ERM) and business continuity management (BCM) activities.

The main innovation of TIMBUS project is therefore its focus on risk assessment based digital preservation of business processes, thus not only bringing together but also advancing traditional digital preservation, risk management and business process management disciplines. Preservation is often considered as a set of activities carried out in isolation within a single domain, without of taking into account the dependencies of third-party services, information and capabilities that will be necessary to validate digital information in the future. Existing DP solutions focus on more simple data objects which are static in nature. The unique aspect of TIMBUS is that it is attempting to advance state of the art by figuring out how more complex digital objects can be preserved and later restored in the same or different environments.

The proposed solution that is therefore based on two premises: Acquisition of business knowledge and risk management.

2.1 Context Acquisition

This step concerns determining, capturing, modelling and indexing business knowledge into an information system, which is a very challenging task. The information model needs to be flexible enough to handle the heterogeneity in concepts and business information, able to comprise information from a business environment consisting in multiple systems, computers and other devices, people and other sources of information.

The proposed approach is to have an open architecture which is able to expand its ability to contact these information sources. In this sense it is able to retrieve information with more or less human intervention (i.e., although it is capable to access the information sources directly, it must be also able to receive information retrieved by other tools or by human intervention, this is relevant to handle some tricky security problems and to enhance confidence on the target systems). The proposed method is to develop generic extraction mechanisms, ones which are able to attach numerous specifically implemented extractors or adapters, which then are able to retrieve correctly the information.

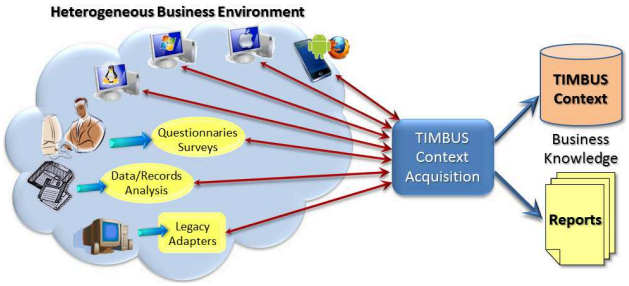

Figure 1: Acquisition of business knowledge metadata

Figure 1 shows the overall architecture of the Context Acquisition module, which works in a multi-agent system paradigm, with dedicated and tailored agents extracting business metadata from the target environment. A central acquisition module gathers the inputs from the multiple agents running on heterogeneous operating systems, environments and platforms, legacy systems and records and business tacit knowledge captured by questionnaires and interviews, reasons these inputs, and models and stores the data into the information system.

The proposed organisation of the stored metadata is to store concepts and relationships into an ontology that properly consolidates this information, supported by a database to persist the remaining specific data. The ontology schema for the central information system (known as Business Information Context) was defined and tailored in the ArchiMate® [1] modelling language, which was considered the best to capture all relevant aspects of an organization and its IT systems. The remaining data is derived from the concepts defined on this ontology, and is shaped into a proposed standard format "Common Upgradeability Description Format" (CUDF) [6], considered the most proper for storing the data and relating it to the ontology concepts.

Business Context metadata is therefore captured by multiple sources (e.g., software agents, surveys and interviews, databases). However, this operation needs to reflect the continuous changes in the business environment, hence besides baselining the information capture, a module shall periodically (or on-demand) access the business environment, comparing the current state with the baselined metadata, and updating it into subsequent baselines.

2.2 Context Analysis and Risk Management

The captured business information *per se* is an important asset towards the objective of being able to restore knowledge, but is clearly not enough in the pursuit of having business continuity goals. To achieve this purpose, the captured and organised information needs to be subsequently analysed to determine risks related to the business.

The intelligent Enterprise Risk Management (iERM) is another proposed module, whose purpose is to, by analysing the captured context information, raise risks that may affect the business stability. This module performs a semantic analysis over the context ontology, conducting a semi-automated generation of risks related to the involved business concepts and facts. It also permits the development of business scenarios and simulations, which are then able to model additional risks. Additionally, multi-disciplinary teams analyse the context data for risks.

One of the most important sources for business-impact risks is the legal aspect. Licences, legislation, standards, processes and procedures are some which may cause severe impact to the business continuity. Hence, besides proposing the strategy of

having multi-disciplinary teams to search for problems, recommendations and opportunities, TIMBUS explicitly demonstrates this by including on its consortium teams that are specialised in several areas like software development, hardware, tests, sensors, cloud and big data, and most especially, a team dedicated solely to Legalities and Life cycle issues, which has the task to determine the risks related to this important area. The other specialised teams are able to determine as well additional risks that were not automatically added by the tools. Risk Management may also identify opportunities that enable an improved functionality of the business, as shown on Figure 2.

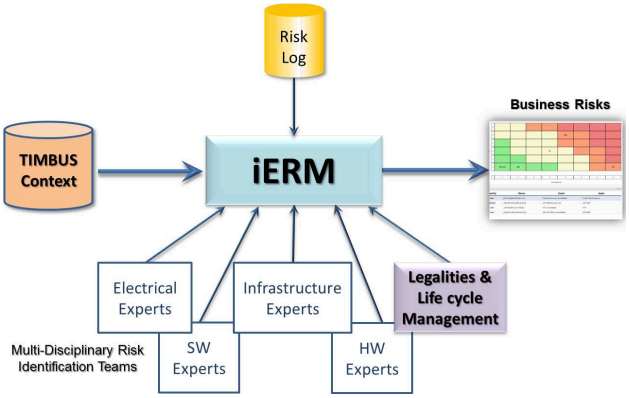

Figure 2: Business Context Analysis

The resulting outcome is a list of risks and opportunities which undergoes the traditional scrutiny of qualification regarding its probability and impact (again, this task is semi-automated by the iERM module, some of the risks are already assigned to a probability and impact, subject to confirmation), to be able to provide a hierarchized and classified list of the most important.

The consequential list of categorised risks follows the risk management methodology steps of determining solutions and mitigation strategies that are able to reduce the probability and impact of the negative risks and increase the same for the positive opportunities. Some of these risks may be considered by risk management as acceptable due to their residual probability or impact, and hence, be considered negligible, others may not.

Several strategies may then be used for this purpose (e.g., insurances, alternative providers, redundancy, infrastructures, software escrow services). Again, TIMBUS advises to have multiple solution providers and exemplifies this by proposing one of these providers – Digital Preservation (Figure 3).

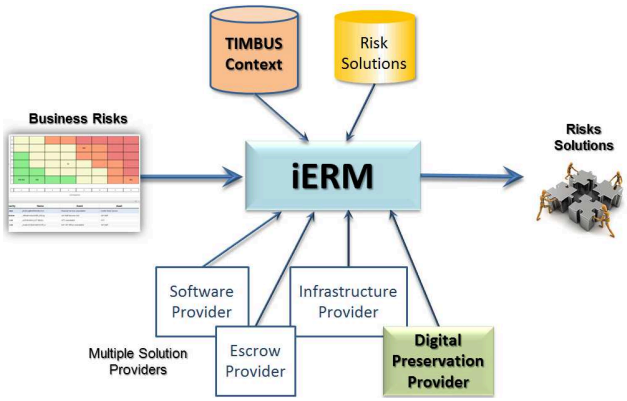

Figure 3: Risk Solution Determination

Each of the solution providers shall receive the risks, analyse the alternative solutions to solve the case, and conclude if it can provide a solution for that risk. In the case of a positive answer, it needs to propose a technical solution and determine the estimated cost for this solution. TIMBUS again provides the advantage of being able to customise and optimise solutions based on the fact that it perceives the needs of the business as this is available in the context information system. This context allows providers to better understand and tailor their solutions to match the business values, vision and objectives. In the particular case of Digital Preservation, this solution provider needs to access the various alternatives for preserving the risk target. These can vary and are very dependent on the types of risk and on the targets, from e.g., environment virtualisation, emulation, to data conversion to other formats, storage in different media, information redundancy. After determining the most appropriate solution for long-term preservation, this solution must undergo a determined cost model to establish its cost estimation.

2.3 Digital Preservation solution

The Digital Preservation module goes beyond simply storing the target artefacts; it fetches the business context to determine all its dependencies, conflicts and surrounding environment, and reasons a proper solution that is able to capture all needed information. This reasoning considers not only the currently stored artefacts in a shell, but also the ones that were previously preserved and baselined (Figure 4). The same procedure is also applicable for the redeployment activities; the alternative analysis determines the best solution for redeploying the preserved artefacts and the reasoning module analyses the new environment and the stored dependencies and constraints to determine a solution for redeployment of the business process environment.

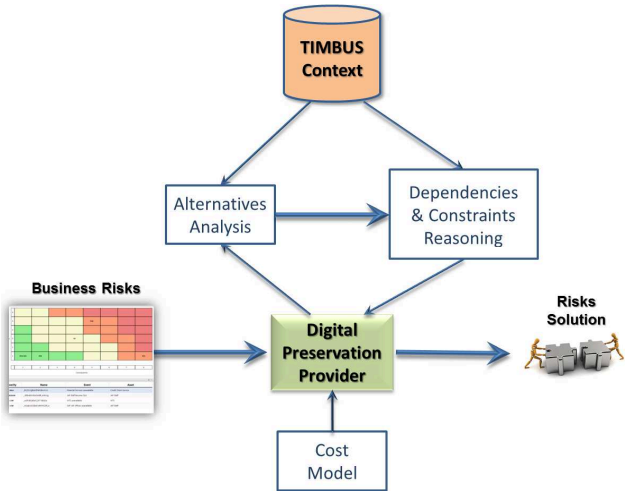

Figure 4: Digital Preservation provider

3. VALIDATION USE-CASES

The TIMBUS project is being validated in three major business areas, with use-case environments and integration with industry:

- The open-source area includes the analysis of repository systems, e.g., EPrints, DSpace, RODA, Alfresco to ensure long-term accessibility of its preserved digital objects and workflows (in cooperation with similar projects SCAPE [8] or the DLM Forum [3]);
- Large-scaled civil engineering scenarios, involving sensors and similar devices, mathematical and data analysis and simulation, and CAD/CAM services, in cooperation in industrial real use-cases with the Portuguese National Civil Engineering authority (LNEC), one of TIMBUS partners;
- Large-scaled multinational use-cases in the health and care support area, with analysis of massive information transactions and decision support workflows in eHealth environment. This is very important, as it touches very important areas such as business secrecy and preventing industrial espionage, as well as large-scaled data consistency and integrity.

4. CONCLUSIONS AND FUTURE WORK

As digital technology is becoming the kernel of the modern economy, supporting different business transactions and providing critical information for making important business decisions, this results in increasing dependencies on refined and complex information systems. Globalisation and its resulting constant evolution paradigm brings new challenges for any industrial organisation to support their business environment, where new hardware and software solutions are developed and adopted on a regular basis. These new versions of components often have limitations when it comes to backward compatibility, which may cause significant downside for modern businesses. This is the underlying cause of most digital preservation issues.

The majority of existing digital preservation solutions focus on long-term storage of digital content such as e.g., documents, images, video, audio and other files. Preservation of an IT infrastructure for supporting business processes much more challenging, requiring the preservation of software and hardware as well as of the related contexts which contain the execution foundations for running processes and workflows. Hence, the problem of digital preservation is well-understood for data-centric information scenarios within a single domain, but not so clear for scenarios where important digital information has to be preserved together with the execution contexts, and with the need to take into account the dependencies of third-party services.

TIMBUS endeavours to preserve both the functional and non-functional specifications of services and software, along with their contexts and dependencies. This type of digital preservation is considerably more complex than only preserving static digital objects and media, as it not only requires dealing with the structural properties of digital information and objects, but also with the dynamic behaviour of processes and their supporting services and software over time, which may include potentially different parties, different enabling technologies, different system components, changed services by different service providers (e.g. SaaS) and differences in other aspects of the contexts of the business processes in focus.

To be able to cope with such dynamic and constantly changing environments, TIMBUS contains several proposals for capturing and indexing the business context, storing and maintain its updates and baselines into a central ontology and data repository. This information is also crucial for all subsequent steps of risk-management-driven business continuity and digital preservation.

Currently, TIMBUS is starting its third year of development, where most research-based activities have been completed and the implementation of several modules is advanced. Some small-scaled software environments are being captured and redeployed as proof-of-concepts, and the methodologies and tools developed in the scope of the project are starting to be applied to the use-cases defined in section 3. The project has just undergone its second year review of proposed deliverables which was approved by the European Commission board of analysis.

5. ACKNOWLEDGMENTS

The author wishes to acknowledge the support of the European Commission through the funding of the TIMBUS FP7 project.

6. REFERENCES

[1] ArchiMate Web Page: *http://www.opengroup.org/subjectareas/enterprise/archimate*. Accessed: 2012-12-14.

[2] Bohannon, J. 2012. DNA: The Ultimate Hard Drive. *Science*.

[3] DLM Forum web page: *http://www.dlmforum.eu*. Accessed: 2013-03-20.

[4] Finnegan, M. 2013. Boeing 787s to create half a terabyte of data per flight, says Virgin Atlantic. *Computer World UK:*

[5] Gantz, J. and Reinsel, D. 2012. *The Digital Universe in 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East Executive Summary : A Universe of Opportunities and Challenges*.

[6] Mancoosi CUDF web page: *http://www.mancoosi.org/cudf*. Accessed: 2013-03-04.

[7] Mearian, L. 2013. Helium-filled hard drives may lift WD to the top. *ComputerWorld*.

[8] SCAPE Project page: *http://www.scape-project.eu*. Accessed: 2013-02-14.

[9] TIMBUS Project page: 2013. *http://timbusproject.net*. Accessed: 2013-05-15