Preservation and Redeployment of Sensor Acquisition Processes from a Dam Safety Information System

Anja Bachmann[1], Martin Alexander Neumann[1], Hossein Miri[1], José Barateiro[2], Gonçalo Antunes[3] and Artur Caetano[3]

[1]*Karlsruhe Institute of Technology (KIT), TECO, Vincenz-Prießnitz-Straße 1, 76131 Karlsruhe, Germany*

[2]*Laboratório Nacional De Engenharia Civil (LNEC), Avenida Brasil 101, 1700 Lisbon, Portugal*

[3]*Instituto de Engenharia de Sistemas e Computadores - Investigação e Desenvolvimento (INESC-ID) , Rua Alves Redol 9, 1000-029 Lisbon, Portugal*

{*bachmann, mneumann, miri*}*@teco.edu, jbarateiro@lnec.pt,* {*goncalo.antunes,artur.caetano*}*@ist.utl.pt*

Keywords: Digital Preservation, Sensor Networks, Civil Engineering

Abstract: There is a growing demand for digital preservation of, not only static objects and files, e.g. text documents and images, but also dynamic objects and whole processes, such as interactive media and entire business and scientific processes. This paper investigates the problem of digitally preserving monitoring processes of a dam safety information system. Monitoring processes are a crucial element in ensuring structural safety; the interpretation of the past data produced by such processes as well as the prediction of potential future behaviour facilitates an earlier detection of possible dam failure. After a successful preservation performance, relevant data can be used or re-produced without the need for the original system to still exist; merely by re-playing the preserved information and data. This enables several possibilities in the scope of a water dam system. The retracing of former situations and structural behaviour decades later is one of them. Furthermore, the interpretation of past data and subsequent prediction of future behaviour that could facilitate an earlier detection of a fault or possible dam failure. This work presents a methodology for preserving the obtained sensor data (readings, measurements, and meta-data) from a dam safety information system, whose involved processes include: data acquisition, the preservation process itself, and the re-playing and redeployment of the preserved data.

1 INTRODUCTION

Within the last few years, there has been a growing demand for digital preservation. This is caused by the need to protect information of enduring value in interrelation with the drastically increasing number of documents and digital objects that are being produced every day (Conway, 1996; Borghoff et al., 2006). This amount has not only to be managed, but also preserved for later use and contextual interpretation, in consideration of the preservation of the value comprised in this data.

Digital preservation research is concerned with providing long-term *access to* and *intelligibility of digital objects*, regardless of their complexity (Lee et al., 2002; Giaretta, 2011). The focus is on preserving digital objects along with their meta-data (contextual or proxy information) required to achieve this goal (Day, 2001). In the past, static objects were considered the main focus of digital preservation. These objects do not expose any behaviour, i.e. they do not perform any activity or interaction with their environment that is visible externally. Such objects, which are to be preserved digitally, are primarily text and multimedia documents. Notably, in order to be generated and interpreted, digital objects require a technological context defined by specific software and in some cases even by specific hardware (Chou et al., 2012).

Recent research in the area of digital preservation has focused on extending established preservation approaches to dynamic objects. These are objects that actually perform behaviour over time. Examples are video games (Guttenbrunner et al., 2010), interactive art (McHugh et al., 2010; Becker et al., 2007) and computational environments, such as scientific workflows (Roure et al., 2011). Additionally, more and more static digital objects are replaced by dynamically generated ones, e.g. generated meta-data and dynamic websites. Digital business processes are a special and more complex type of dynamic object, which present big challenges to preservation due to the multitude of dependencies that have to be kept in mind. As a result, there is a necessity to not only preserve the object but rather the process describing the context of the digital object to fulfil the aim of preserving this object fully.

Here, important civil engineering structures like dams are considered. The monitoring of their behaviour is important to guarantee is faultless function. The data required for analyzing the behaviour is acquired by a set of sensors placed along the dam. To facilitate the monitoring and guarantee its functionality in the future, the monitoring and data acquisition process have to be preserved. It is important to preserve data about the water dams so that later on it is possible to retrace former situations and behaviours. Preserved data can be used for example to build a model and predict future behaviour or to detect anomalies. It also facilitates to track a possible problem source down to specific system behaviour. To preserve data, we need to keep both, sensor data and contextual information, in a preservable format.

The remainder of this paper is structured as follows. Section 2 gives an overview of the background of our use case and the monitoring processes of the water dam. In Section 3, we present a typical monitoring architecture for water dams and name relevant aspects to be preserved. This is followed by our proposed methodology to preserve these contextual aspects in Section 4. This extends our general architecture for preservation of business processes which has been previously published in (Galushka et al., 2012). The preserved data can, then, be re-played for contextual interpretation. Section 5 describes this process in detail. Finally, Section 6 discusses the usefulness and suitability of our architecture and model in the context of preserving and re-playing water dam monitoring processes, and also gives an outlook to future work.

2 MONITORING OF WATER DAM BEHAVIOUR

We perceive a monitoring process of a water dam as such a kind of process to be preserved. In this paper, we present our perspectives on *what*, *why* and *how* it is to be preserved, in the context of such water dam monitoring processes. Our general formal model to capture *what to preserve* has been presented in (Neumann et al., 2012).

Here, we present our specialization of this model to the domain of water dam monitoring processes, and will discuss its usefulness in preserving these monitoring processes.

2.1 Use Case Scenario

Within the last years, the interest in the area of civil engineering raised. The focus is on “conservation of the environment, the welfare and safety of the individual” and there is a need to “manage the prevailing natural and manmade risks in a conscious, consistent and rational manner” (Faber, 2007). Wieland and Mueller (Wieland and Mueller, 2009) advise their readers of the far-reaching consequences of a possible failure of large civil engineering structures. Especially in systems like dams or nuclear power plants a failure could endanger the life of human beings and the nature, but also damage important infrastructures like electricity or water supply. There are several preventive measures and mitigation strategies to reduce the risk. An often applied methodology is to detect anomalies and errors as early as possible to increase the time to react to them and prevent from the worst consequences.

In this use case, the focus will be on large civil engineering structures like dams. They have a critical infrastructure that can face the risks mentioned above. Some countries established legal obligations for national research institutions to monitor the behaviour of a civil engineering system. Portugal for example enforced a strict Dam Safety Legislation which are stating some general regulations for water dams, among others. This legislation also instructed the Laboratório Nacional De Engenharia Civil (LNEC)[1] to monitor the structural safety of a water dam system. The system, considered here, encompasses 117 water dams in Portugal. Each of them is equipped with a different number of sensors, most with less than 50, some even with more than 1000. Overall there are 31479 sensors that are taken into account.

To preserve all data relevant for the preservation process in the context of this use case, it is necessary to gather information about the sensor, its environment and its measurements (Kutter et al., 2002). Such information has to be exchanged between several programs, e.g. the data collector and the evaluation system. Hence, it is necessary to select a suitable and interoperable file format to perform this task. The resulting tasks are therefore to acquire process information, sensor data and environmental information, convert the data into a preservable format and, whenever required by the user, replay this data for contextual interpretation.

To perform this task it is necessary to install several sensors in the dam structure. It is possible to

[1] http://www.lnec.pt/

manually or automatically gather the required information and later on convert this data for further analysis. Data is acquired manually, e.g. by a human operator using measuring devices at different locations within the water dam. For automatic data collection it is advisable to establish a wireless network to directly get the data from the sensor and store it. If such a connection is established, the sensor can communicate directly and automatically with the dam safety information system and transfer the information.

2.2 The Need for Preservation

Monitoring a dam and understanding its structure are inseparable from each other. It is possible to evaluate and validate measurements, taken from sensors installed at strategic points within the observed system, and interpret correlations between them. The resulting information can further be used for the prediction of future behaviour, the estimation of missing values or the detection of anomalies. This underlies the requirement of reliable data and measurements. Devices like plumb lines and piezometers are installed at different locations within a water dam to gather information about it and monitor its behaviour (American Society of Civil Engineers, 2000; International Commission on Large Dams, 1999). The number of used sensors can range from less than 100 to more than 1000, depending on the dimension of the dam.

Process information, meta-data and sensor measurement – all this data has to be preserved. In this way, contextual interpretation is possible even decades after the data was gathered and even without the existence of the original system. Occurring problems are that the former processes are not existing any more and cannot be retraced or that former data is missing and cannot be reproduced easily. Preservation countervails these problems.

Altogether, the motivational factors for digital preservation of sensor data acquisition processes in the scope of civil engineering can be summarized as follows.

- *Compliance with legal requirements* – Fulfilling the legal obligations and the agreements established with the owners of the structures.
- *Replay of processes* – Enabling a later replay of former scenarios for retracing the scenario and the original behaviour of the underlying system. Ensuring the authenticity, correctness and validation of the execution.
- *Re-use of processes* – Gaining new insights by modifying parts of the execution process. Re-execution of former processes with modified parameters or execution of a modified process with the original data.
- *Assessment of the costs of retention* – Considering whether to keep a big dataset of gathered data or to only keep the process that generate the data with respect to the corresponding costs.

In consideration of these factors, the preservation of processes in the scope of civil engineering monitoring activities is a challenge that must be addressed.

3 DATA ACQUISITION

In digital preservation it is not sufficient to only store measurements that were gathered by sensors. For a later understanding of a situation or a scenario it is inevitable to also store information about the process and meta-data about the environment and interrelation of sensors. This environment encompasses the location of the sensor (position, location), but also environmental information like weather, temperature or humidity. Additionally, there is also contextual information about the process itself, e.g. dependencies among the software or hardware components.

The focus of this project is on process preservation. It is inevitable to preserve all this data which incorporates information about the process or represents influencing factors. Environmental components (temperature, location, position) influence the functionality of a sensor, the measuring and monitoring processes. Dependencies and interrelations of software or hardware components are relevant to model the process.

3.1 Acquisition of Sensor Data

As mentioned before, we require sensor data and meta-data, among others. The sensors have to be accessed and the measurements have to be gathered and afterwards stored in a database. This data consists of the identification number of the sensor, the identification number of the measuring device, the measured value, information about the temperature and so on. It is possible to acquire data by manual or automatic measurement. For example it is possible to automatically get location information by checking the GPS signal. Values like temperature can also be measured automatically. Information about the monitoring process itself or dependencies will have to be added semi-automatically in the next step.

Currently, the Laboratório Nacional De Engenharia Civil is handling 35 types of manual sensors and 25 types of automatic data gathering devices. The processes of gathering and transforming the measurements and all involved data transformations and algorithms are what has to be preserved within this use case.

3.2 Water Dam Context Model

Several aspects in the context of a business process have to be taken into account for digital preservation to ensure successful redeployment of that process (Neumann et al., 2012). The authors also define two kinds of aspects that have to be taken into account for the preservation of a process: *coarse-granular aspects* (abstract and not only relevant for the process but its entire domain) and *fine-granular aspects* (more specific to the process itself and its sub-processes).

If such knowledge about a process is machine-interpretable formalized and stored directly with the digital objects, semantic reasoners can immediately incorporate it into decision making. There is no need to consult external systems, repositories or human preservation operators. Furthermore, if the formal system is not a specialized language, but a generic knowledge representation language such as OWL (Web Ontology Language), the preservation-relevant advantage arises that the knowledge kept close to the digital objects may even be object-specific. This means that the model may be specific to a preserved business process, but still is machine-interpretable (Neumann et al., 2012).

There are two Context Model instances (DIO and DSO) related to this use case which will only be shortly described here. A domain-independent ontology (DIO) represents a neutral, domain-independent language that is able to represent the core concepts of our Context Model. It is designated domain-independent since it does not address any specific domain-dependent concerns. A domain-specific ontology (DSO) represents a domain-specific language that addresses a particular set of concerns. Our Context Model will comprise a core DIO and a set of DSOs. They represent object classes and dependencies and interrelations between them in order to describe the process to be preserved and its context. Contextual information and the dependencies between processes are extracted by using different extractors. One of the DSOs is related to this use case – the Sensor DSO. It aims at capturing particular aspects about sensors and other related information. It gives an overview of the relations between acquisition algorithms and transformations and the sensors, their types and properties. The UML class diagram in Fig. 1 describes the different classes of the Sensor DSO.

4 DATA PRESERVATION

The focus of this project is on digital process preservation. To achieve this the subsequently described methodology to preserve civil engineering processes was performed. As explained in Section 3 the first step is to acquire data. This consists of the acquisition of measurements, but also (meta) information about the sensors and the associated processes. To model a process it is also important to detect and cover dependencies between data and sub-processes. Such contextual information is added to the content data and stored together for preservation issues.

In todays IT domain there is a burden of coping with the diversity and heterogeneity of data sources, dynamically selecting appropriate data sources, and scaling of data from mobile, distributed data sources (Cohen et al., 2001). This causes the need for an interoperable, easily manageable, flexibly extensible and common data format that facilitates the access of and further work on a document (Bloechle et al., 2006). An interesting candidate is XML with a sensor data extension, e.g. Unisens[2], SenML[3] or SensorML[4]. They provide various features relevant to our use case.

The presented preservation approach divides the set of required data into three parts: sensor measurements, sensor meta-data and process meta-data. A separation of these components was chosen to allow working on one component without influencing or endangering the other ones. Two different data file types are used to preserve this data. The sensor measurements and the environmental sensor meta-data (e.g. temperature) are stored in one of the mentioned XML-based formats. The process meta-data (e.g. dependencies and interrelations of processes or sensors) is stored in the Context Model, introduced in Subsection 3.2. A specific part of the Context Model, the OWL file "SensorDSO", is capturing dependencies and properties of sensors. All entities, dependencies and properties of the DSO can be mapped to the acquired sensor data and meta-data. The advantage: OWL is not only interoperable, it also allows a reasoning on the data to induces findings about dependencies and properties. Thereby, the incorporated meta-data will be contextual interpretable and human understandable.

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

[3]http://tools.ietf.org/html/draft-jennings-senml-08

[4]http://www.opengeospatial.org/standards/sensorml

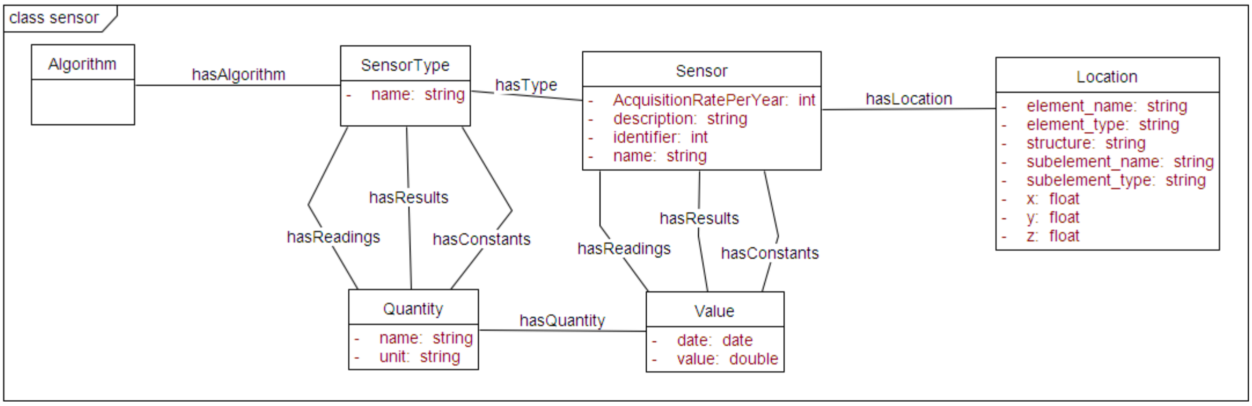

Figure 1: UML Class diagram describing the classes of the Sensor DSO.

5 DATA REPLAY

One of the main purposes of the digital preservation of business process is to enable a later contextual interpretation of a former scenario and to retrace what has happened before. For realizing this, preserved data is replayed. As stated by van der Hoeven et al. the focus is not only "on the digital object, but on the hardware and software environment in which the object is rendered", moreover the aim is to "(re)creating the environment in which the digital object was originally created" (van der Hoeven et al., 2007). In our case, this means to preserve and replay the monitoring process.

The replay of a former scenario is important to retrace what had happened before or to make further use of this historic data. There are several alternatives to perform an emulation of past sensor data.

- One method: as input data, a certain time and date combination and the ID of a sensor are given. Afterwards it is possible to track down this one sensor to the time back then and retrieve the past measurement.
- A second way: again, time and date for the emulation are given. This time, a whole water dam/sensor system is considered and not a single sensor. Then, by using the past measurements and the contextual information, it is possible to get an overview of the former behaviour of this water dam.
- It is also possible to apply machine learning techniques to the sensor data and contextual information. A model is trained and learned. Later on this model is able to reproduce information about former behaviour based on given input.

Based on the concrete application and its complexity, a different method has to be chosen. Probably, there are even more possible replaying methods, the presented ones are only giving a first insight. A preserved process has been redeployed successfully if the replayed process shows the same behaviour as the original one.

It is not only possible to use the preserved data to replay former scenarios and interpret them. The data can also be used as historic input data for several stream mining algorithms. It is possible to predict the future behaviour of a water dam based on this data. Moreover it is possible to learn a model to detect anomalies. It is also possible to compare different scenarios.

6 CONCLUSION AND OUTLOOK

The relevance of digital preservation in areas such as civil engineering is rising due to the evolvement of big data and the increasing importance of safety civil infrastructures. This paper underlined the significance of digital preservation of monitoring such an infrastructures. It further presented a methodology for preserving complex dynamic processes of a water dam system. Still, there are open tasks and starting points for further investigations.

Digital preservation is challenging. Its aim is not only to store cultural heritage but also to enable contextual interpretation and retracing of the past. We are investigating a use case situated in the area of civil engineering. One aim of civil engineering is to increase safety and reduce the risks of a failure by a better monitoring and by future anomaly detection. Here, the focus is on one step before: the preservation of data gathering and monitoring processes, containing information of what already happened and is happening. We presented a methodology for this preservation, bringing civil engineering one first step closer to risk mitigation and failure avoidance.

The considerations presented in this paper are a good basis, but requesting further research. First of all it is inevitable to develop a conversion tool for preservation. This tool is supposed to take sensor and contextual information as input and convert them into a preservable format. Of course, it is necessary to check whether the preserved data contains all relevant information that was given by the input data. The preserved data will be replayed some time later on. To perform this action a replaying algorithm is required. This algorithm has to be examined regarding its replay accuracy and usefulness. It is important to check whether the replayed data behaves like the original data did. As mentioned before in Subsection 4 there are additional possibilities to work with preserved data in the future than just replay it. There is also the option to use the preserved, historic data to predict future behaviour, e.g. by performing stream mining algorithms. Especially in the framework of civil engineering, where the focus is often on enhancing the safety and reducing a risk by avoiding failures, the early detection errors or uncommon behaviour is important. This can be achieved for example by learning a model by applying machine learning techniques on the data. It is considerable to simulate emergency situations, e.g. the bursting or the failure of a water dam, by altering some parameters. The task for the developed system is to detect this anomaly as soon as possible to raise an alarm. The question is whether the system is able to detect such a failure in time.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the funding by the European Commission under the ICT project "TIMBUS" (Project No. 269940, FP7-ICT-2009-6) within the 7th Framework Programme. This work was also supported by national funds through FCT Fundação para a Ciência e a Tecnologia, under project PEst-OE/EEI/LA0021/2013.

REFERENCES

American Society of Civil Engineers (2000). *Guidelines for instrumentation and measurements for monitoring dam performance*.

Becker, C., Kolar, G., Kueng, J., and Rauber, A. (2007). Preserving Interactive Multimedia Art: A Case Study in Preservation Planning. In *Proceedings of the 10th International Conference on Asian Digital Libraries*, ICADL'07, pages 257–266, Berlin, Heidelberg. Springer-Verlag.

Bloechle, J. L., Rigamonti, M., Hadjar, K., Lalanne, D., and Ingolf, R. (2006). XCDF: A Canonical and Structured Data Format. *Document Analysis Systems VII*, 3972:141–152.

Borghoff, U. M., Rödig, P., and Scheffczyk, J. (2006). *Long-Term Preservation of Digital Documents, Principles and Practices*. Springer Berlin Heidelberg.

Chou, C., Dappert, A., Delve, J., and Peyrard, S. (2012). Describing Digital Object Environments in PREMIS. In *Proceedings of the 9th International Conference on Preservation of Digital Objects*, iPRES 2012.

Cohen, N., Ourakayashta, A., Turek, J., Wong, L., and Yeh, D. (2001). Challenges in Flexible Aggregation of Pervasive Data.

Conway, P. (1996). Archival Preservation in a Nationwide Context. 2:204–222.

Day, M. (2001). Metadata for Digital Preservation: A Review of Recent Developments. In *Proceedings of the 5th European Conference on Research and Advanced Technology for Digital Libraries*, ECDL '01, pages 161–172, London, UK, UK. Springer-Verlag.

Faber, M. H. (2007). Risk and safety in civil engineering.

Galushka, M., Taylor, P., Gilani, W., Thomson, J., Strodl, S., and Neumann, M. A. (2012). Digital Preservation Of Business Processes with TIMBUS Architecture. In *Proceedings of the 9th International Conference on Preservation of Digital Objects*, iPRES 2012.

Giaretta, D. (2011). Preservation of intelligibilityintelligibility of digital objects. pages 39–166.

Guttenbrunner, M., Becker, C., and Rauber, A. (2010). Keeping the Game Alive: Evaluating Strategies for the Preservation of Console Video Games. *International Journal of Digital Curation*, 5(1).

International Commission on Large Dams (1999). *Guidelines for automated dam monitoring systems*.

Kutter, B., Wilson, D., and Bardet, J. (2002). Metadata Structure for Geotechnical Physical Models (and Simulations?). Technical report.

Lee, K.-H., Slattery, O., Lu, R., Tang, X., and McCrary, V. (2002). The state of the art and practice in digital preservation. *Journal of Research-National Institute of Standards and Technology*, 107(1):93–106.

McHugh, A., Konstantelos, L., and Barr, M. (2010). Reflections on Preserving the State of New Media Art. In *Proceedings of the Archiving Conference*.

Neumann, M. A., Miri, H., Thomson, J., Antunes, G., Mayer, R., and Beigl, M. (2012). Towards a Decision Support Architecture for Digital Preservation of Business Processes. In *Proceedings of the 9th International Conference on Preservation of Digital Objects*, iPRES 2012.

Roure, D. D., Belhajjame, K., Missier, P., Gmez-Prez, J. M., Palma, R., Ruiz, J. E., Hettne, K., Roos, M., Klyne, G., and Goble, C. (2011). Towards the Preservation of Scientific Workflows. In *Proceedings of the 8th International Conference on Preservation of Digital Objects*, iPRES 2011.

van der Hoeven, J., Lohman, B., and Verdegem, R. (2007). Emulation for Digital Preservation in Practice: The Results. 2:123–132.

Wieland, M. and Mueller, R. (2009). *Dam safety, emergency action plans and water alarm systems*.