# A systematic approach for the assessment of digital preservation activities

Diogo Proença
IST/INESC-ID
Rua Alves Redol, 9
1000-029 Lisbon
diogo.proenca@ist.utl.pt

## ABSTRACT

Assessing and improving an organization's digital preservation activities can rather troublesome and subjective. Different organizations have different levels of detail regarding the planning and operation of their digital preservation activities. Previous work on the assessment of digital preservation activities revealed that the assessment based on checklists and interviews wield subjective, non-repeatable and non-comparable results. As a result, in this paper we present a kernel meta-model which aims at providing a systematic approach for the assessment of digital preservation activities.

## Categories and Subject Descriptors

H.1 [**Information Systems**]: Models and Principles; J.1 Administrative Data Processing Government; K.6.4 Management of computing and Information Systems.

## General Terms

Measurement, Performance, Design, Verification.

## Keywords

Digital Preservation, Assessment, Situational Method Engineering.

## 1. INTRODUCTION

In order to assess and improve digital preservation activities there is the need to have a systematic, repeatable method that enables the assessment of organizations from different domains, with different sizes and different goals. In the past we used checklists to assess organizations, and we applied these checklists through interviews and be getting evidence to make a sound foundation to our scores. However, we realized that different assessors got to different results with the same organization and evidence.

To solve this problem we merged the concepts of situational method engineering (SME), with those of process assessment and improvement, and digital preservation with the aim of creating a kernel meta-model which will create an abstraction of an organization and enables the systematic assessment of organizations.

This means that in order to assess an organization we extract features from the organization using various techniques, such as, extraction/elicitation techniques, process mining, questionnaires, or the Delphi method. Then, we populate the meta-model using the information extracted. Finally, we reason on the organization model using a set of predefined questions (which are appropriate for the organization).

In order for the meta-model to be applicable to different types of organizations it has to follow some principles, as described below.

1. Concern-orientation. The meta-model shall represent the concepts necessary and sufficient to address an explicit set of modelling concerns. This means that the model shall be derived from the questions that need to be addressed and to provide answers to those questions. This also means that the model shall not support any concepts that are not explicitly derived from concern. The principle of concern-orientation and the principle of viewpoint-orientation are described in detail in the ISO 42010:2011 standard [3]. This standard defines requirements on the description of systems and enterprise architecture.
2. Expressiveness. The meta-model shall be able to represent the domain concepts without ambiguity. This entails defining the minimum set of types and relationships to describe a domain.
3. Extensibility. The meta-model must cope with extensions because digital preservation activities use multiple concurrent perspectives on the same problem. This derives from being able to answer to multiple concerns. Therefore, domain-specific and domain-independent models must coexist and the overall meta-model must cope with multiple model transformation and integration.
4. Viewpoint-orientation. The model must support defining views over subsets of its concepts. This serves to facilitate the communication and the management of the models as viewpoints act as a separation of concerns mechanism. Viewpoints will facilitate addressing multiple concerns and managing the multiple extensions required to handle these concerns.
5. Modularity. The models must follow the principles of high-cohesion and low-coupling. Observing these principles contributes to expressiveness and extensibility of the meta-model. It is especially important that adding new domain-specific aspects to the meta-model does not interfere with the concepts and relations already present in the kernel meta-model.

With all these principles in mind, we present the SME domain in Section 2, then we present the kernel meta-model in section 3. In section 4 we depict how to extend the kernel. Finally, in section 6 we draw conclusions and provide sight on future work.

## 2. SITUATIONAL METHOD ENGINEERING

Situational Method Engineering is a branch of method engineering (ME), ME is "the engineering discipline to design, construct and adapt methods, techniques and tools for the development of information systems" [4]

There are several research areas in ME from which we present four; (1) Meta-modelling techniques, which are special purpose specification techniques for the design and evaluation of methods and tools [5]; (2) Tool Interoperability, there are lots of tools that cover only part of the method development life cycle and which integration is of great value; (3) Situational Methods, because all projects are different there cannot be a single standard method this is were situational methods come into play and; (4) Comparative review of methods and tools, quality of methods comprise aspects, such as, completeness, expressiveness, understandability, effectiveness of resources, and efficiency [6][7][8].

In the following sections of this paper we will focus on point (3) Situational Methods. Situational Methods is an information systems development method tuned to a specific situation/context [9]. There have been several authors who have recognized the importance of situational methods, such as, [9] to [14].

There are several ways to represent the knowledge in the SME field as for example:

1. Fragment Based, which provides components that are appropriate to a certain context. The created fragments are supported by component assembly rules and constraints which must be satisfied by the overall created method;
2. Chunk based, works by associating reusable components to their description which facilitates components research and extraction as specified by the user's needs;
3. Pattern based, a pattern describes a recurring problem with its associated solution.

In this paper we will focus on the fragment based SME methods.

## 3. A KERNEL META-MODEL FOR DIGITAL PRESERVATION ASSESSMENT

The kernel meta-model contains the set of entities and relationships essential to describe and assess the digital preservation processes as well as to assess the process maturity. Each of the processes should be mapped to the kernel to enable the assessment. The motivation behind this kernel meta-model is to create an abstraction of an organization so that we can apply different assessment methods or different assessment method fragments as advocated by situational method engineering. The reason for applying different methods or fragments is due to fact that different organizations have different sizes, different processes and different level of formal specification of their processes and activities. Therefore, there is the need to assess different types of organization in different ways using the appropriate methods or fragments.

In Figure 1 we present the actual kernel meta-model, in Table 1 we detail the concepts and in Table 2 we detail the relations between the concepts.

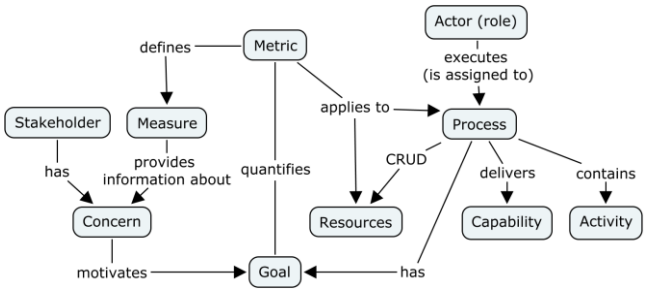


**Figure 1 - Kernel Meta-model**

**Table 1 - Kernel Meta-model Concepts**

| Name | Description |
|---|---|
| Activity | Work that is performed within a business process. An activity can be atomic or non-atomic (compound). There are three types of activity that are part of a Business Process: Process, Sub-process, and Task [2] |
| Actor (Role) | Definition of the responsibilities for performing business processes. There are various types of roles, such as: Customer, Supplier, and Technology. [2] |
| Capability | An ability that an organization, person, or system possesses. Capabilities are typically expressed in general and high-level terms and typically require a combination of organization, people, processes, and technology to achieve. For example, marketing, customer contact, or outbound telemarketing. [1] In many cases a capability will be seen as a synonym for business service as defined in enterprise architecture. |
| Concern | Key interests that are crucially important to the stakeholders in a system, and determine the acceptability of the system. Concerns may pertain to any aspect of the system's functioning, development, or operation, including considerations such as performance, reliability, security, distribution, and evolvability. [1] |
| Goal | A statement about a state or condition of the enterprise to be brought about or sustained through appropriate Means. A Goal amplifies a Vision; that is, it indicates what must be satisfied on a continuing basis to effectively attain the Vision. A Goal should be narrow, focused enough that it can be quantified by objectives. [2] |
| Measure | An indicator or factor that can be tracked, usually on an ongoing basis, to determine success or alignment with objectives and goals. [1] |
| Metric | Measures of performance are defined in an enterprise's Business Motivation Model as objectives. They may be based on risks and potential rewards identified in assessments. [2] |
| Process | A set of activities that are performed within an organization or across organizations. A business process may contain more than one separate process. Each process may have its own sub- |

| | |
|---|---|
| | processes. Individual processes would be independent in terms of sequence flow, but could have message flows connecting them. [2] |
| Resources | An asset that is consumed in the operations of the enterprise and replenished. Examples: Raw materials, parts, Finished goods, cash. Some resources are capacity of fixed assets over time; for example production capacity, storage space, people's time – they are consumed, or dissipated by not being used. [2] |
| Stakeholder | An individual, team, or organization (or classes thereof) with interests in, or concerns relative to, the outcome of the architecture. Different stakeholders with different roles will have different concerns. [1] |

**Table 2 - Kernel Meta-model Relations**

| From | Relation | To | Rationale |
|---|---|---|---|
| Process | Contains | Activity | A Process contains a set of criteria. |
| Process | Delivers | Capability | A Process delivers an organizational capability. |
| Process | CRUD | Resources | A Process creates, uses and depletes the resources used by it. |
| Actor (Role) | Executes (is assigned to) | Process | A Process contains several actors and roles assigned to it; these actors maintain the execution of the process. |
| Process | Has | Goal | The process enables that a set of goals are attained. |
| Metric | Applies to | Process | A process contains a set of metrics that are used to assess the maturity, effectiveness or efficiency of the process. |
| Metric | Quantifies | Goal | The metrics are used to quantify the goals. |
| Metric | Defines | Measure | The metrics define measures (units of measure) for the measurement of the metrics. |
| Concern | Motivates | Goal | A concern raised by a stakeholder motivates a goal. |
| Measure | Provides information about | Concern | The measures provide useful information about the treatment and state of a certain concern. |
| Stakeholder | Has | Concern | A stakeholder raises several concerns regarding various organizational aspects. |

## 4. EXTENDING THE KERNEL META-MODEL

When there is the need to introduce a new aspect to the kernel we might need to extend the kernel meta-model in order to accommodate the new concepts and relations.

For example, to introduce the concepts of Preservation Monitoring, which is the ability to monitor preservation operations and external influencers of interest for certain properties, verify the compliance of operations, and report deviations to preservation planning and preservation context and management, in order to ensure preservation compliance to business requirements. In order to introduce the preservation monitoring concepts we need to identify and map them to the kernel meta-model as depicted in Figure 2, and specify the concepts and relations depicted in Table 3 and Table 4.

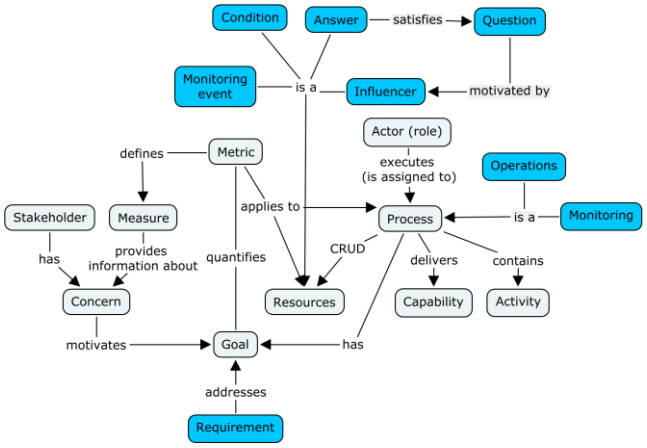


**Figure 2 - Preservation Monitoring extension of the kernel meta-model**

**Table 3 - Preservation monitoring extension concepts**

| Name | Description |
|---|---|
| Condition | A premise upon which the fulfillment of an agreement depends |
| Answer | Monitoring provides answers to preservation planning by providing information on how preservation plans are executed. |
| Question | Planning provides questions which should be answered by the monitoring event of preservation activities. |
| Monitoring event | The process of collecting, analyzing, and signaling event occurrences. |
| Influencer | Something that can cause changes that affect the enterprise in its employment of its Means or |

| | |
|---|---|
| | achievement of its Ends [2] |
| Operations | Control deployment of preservation plans. This includes analyzing content, executing preservation actions and ensuring adequate levels of provenance, handling preservation metadata, conducting Quality assurance, and providing reports and statistics, all according to preservation plans. |
| Monitoring | Ensure that the state of the world and the state of the system are known so that alignment to goals can be verified at any point in time. |
| Requirement | High-level statements of the goals, objectives, or needs of an organization. They usually describe opportunities that an organization wants to be realized or problems that they want to be solved. Often stated in a business case. [16] |

**Table 4 - Preservation monitoring extension relations**

| From | Relation | To | Rationale |
|---|---|---|---|
| Monitoring event | Is a | Resources | A monitoring event is an asset that is used by monitoring to effectively monitor preservation activities. |
| Condition | Is a | Resources | A condition is an asset that is used by monitoring to effectively monitor preservation activities. |
| Answer | Is a | Resources | An answer is an asset that is used by monitoring to effectively monitor preservation activities. |
| Influencer | Is a | Resources | An influencer is an asset that is used by monitoring to effectively monitor preservation activities. |
| Answer | Satisfies | Question | An answer should provide the information necessary to satisfy a question motivated by an influencer. |
| Question | Motivated by | Influencer | Monitoring must provide answers to the influencer's questions. |
| Operations | Is a | Process | Preservation operations is a process that enables planning of the preservation activities |
| Monitoring | Is a | Process | Preservation monitoring is a process that enables monitoring of the preservation activities |
| Requirement | Addresses | Goal | Goals must drive the fulfillment of requirements motivated by the goals. |

## 5. CONCLUSIONS AND FUTURE WORK

By addressing the assessment and improvement of preservation activities using a systematic approach will provide valuable and repeatable assessment results which can then be used to provide usable guidelines to improve current activities.

In this paper, we present a kernel meta-model that provides the minimum set of concepts and relations necessary to assess digital preservation in organizations. Moreover, it can be extended to cover different aspects of digital preservation, such as, preservation monitoring.

This kernel meta-model will then be used in an assessment method, which will be in line with SME guidelines in order to be applicable in the most disparate environments. Using the guidelines provided by the ISO15504 [15] depicted in Figure 3 we include the kernel meta-model in the "Initiate Assessment" phase.

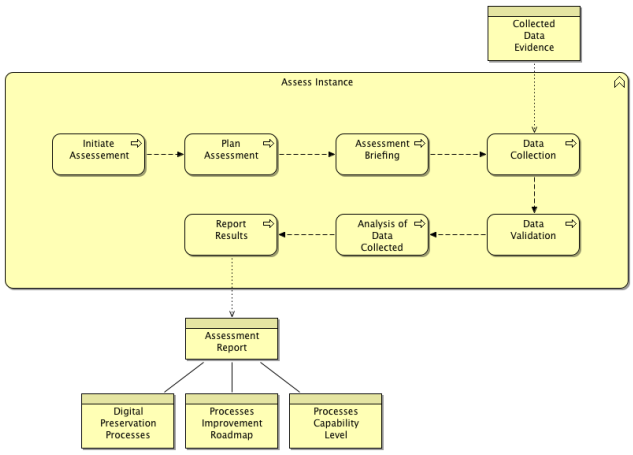


**Figure 3 - Instance Assessment Process**

Drilling down the "Initiate Assessment" phase we can clearly identify where the kernel meta-model will be used as depicted in Figure 4.

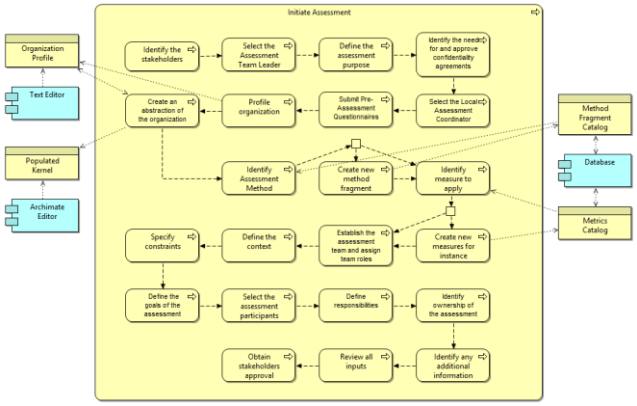


**Figure 4 - Initiate Assessment Method Fragment**

In the "Initiate Assessment" phase, the kernel meta-model will be realized after the organization is profiled and will be used to provide an abstraction of the organization in order to continue the assessment. The populated kernel can then be used to select the most appropriate assessment method fragment for the organization.

As there are various types of organizations we choose to build a method fragment catalog which contains several assessment methods, such as, checklists, interviews, Delphi methods that can be used for different types of organizations regarding their level of process specification, organizations size and business specification. This method fragment catalog will be populated incrementally based on concrete use cases and lessons learned.

The same approach will be used for the measuring the metrics as different organizations have different details and document regarding their activities. One example is the size of certain organization's digital content collection. Some organizations use physical quantities (i.e. Gigabyte, Terabyte) while others use the number of objects in the collection.

In this way we provide a method that can used in the most diverse organizational environments, that is neither oversimplified or generic nor very specific and useful for most organizations assess and improve their preservation activities.

As future work we will consider exporting the kernel meta-model to Web Ontology Language (OWL)[1] and use protégé[2] and SPARQL[3] queries to automatically get answers from the meta-model in order to further standardize the assessment process and provide comparable results regarding the current state of the assessment activities in different organizations as well as a way to benchmark organizations regarding digital preservation.

## 6. ACKNOWLEDGMENTS

Part of this work has been funded by the Vienna Science and Technology Fund (WWTF) through project BenchmarkDP (ICT12-046).

[1] http://www.w3.org/TR/owl-ref/

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

[3] http://www.w3.org/TR/sparql11-overview/

## 7. REFERENCES

[1] The Open Group. *TOGAF Version 9.* Van Haren Publishing. 2009.

[2] Object Management Group. *Business Motivation Model Version 1.1*. 2010.

[3] ISO/IEC/IEEE 42010:2011. Systems and software engineering – Architecture description. 2011.

[4] Ralyte J., Brinkkemper S., Henderson-Sellers B. *Situational Method Engineering: Fundamentals and Experiences.* Proceedings of the IFIP WG 8.1 Working Conference, 12–14 September 2007, Geneva, Switzerland. Springer US. 2007

[5] Brinkkemper, S. *Method Engineering: Engineering of Information Systems Development Methods and Tools.* Information and Software Technology, 38 (4). pp. 275-280. ISSN 0950-5849. 1996.

[6] Song X., Osterweil L. *Toward objective, systematic design-method comparisons*. IEEE Software. 1992.

[7] Hong S., Goor G., Brinkkemper S. *A comparison of object-oriented analysis and design methodologies*. In Proceedings of the 26th Hawaiian Conference on System Sciences (HICSS-26). IEEE Computer Science Press, Vol. IV, pp. 689-698. 1993.

[8] Marttiin P., Rossi M., Tahvainanen V-P., Lyytinen, K. *A comparative review of CASE shells - A preliminary framework and research outcomes*. Information and Management. 25 (1), pp. 11-31. 1993.

[9] Harmsen F., Brinkkemper S., Oei H. *Situational method engineering for information system project approaches*. In Proceedings of the IFZP WG8.1 Working Conference CRIS'94 Maastricht, Holland, Amsterdam, pp. 169-194. 1994.

[10] Kumar K., Welke R. *Methodology engineering: A proposal for situation-specific methodology construction*. Challenges and Strategies for Research in Systems Development. John Wiley. 1992.

[11] Olle T., Hagelstein J., MacDonald I., Rolland C., Sol H., Assche F., Verrijn-Stuart A. *Information Systems Methodologies - A Framework for Understanding*. Addison-Wesley. 1991.

[12] Heym M., Osterle H. *Computer-aided methodology engineering*. 1993.

[13] Hidding G., Freund G., Joseph J. *Modeling large processes with task packages*. Workshop on Modeling in the Large, AAAI Conference, Washington DC. 1993.

[14] Slooten K, Brinkkemper S. *A method engineering approach to information systems development*. In Proceedings of the IFIP WG8.1 Conference on Information Systems Development. 1993.

[15] ISO/IEC 15504:2004. *Information technology - Process assessment.* International Organization for Standardization and International Electrotechnical Commission Std. 2004.

[16] International Institute of Business Analysis. *A Guide to the Business Analysis Body of Knowledge.* 2009.