Ontology-based Representation of Context of Use in Digital Preservation

Efstratios Kontopoulos[1], Marina Riga[1], Panagiotis Mitzias[1], Stelios Andreadis[1], Thanos G. Stavropoulos[1], Nikolaos Lagos[2], Jean-Yves Vion-Dury[2], Georgios Meditskos[1], Patricia Falcão[3], Pip Laurenson[3], Ioannis Kompatsiaris[1]

[1] Information Technologies Institute, Thessaloniki, Greece
{skontopo, mriga, pmitzias, andreadisst, athstavr, gmeditsk, ikom}@iti.gr
[2] Xerox Research Centre Europe (XRCE), Meylan, France
{Jean-Yves.Vion-Dury, Nikolaos.Lagos}@xrce.xerox.com
[3] Tate, London, UK
{Patricia.Falcao, Pip.Laurenson}@tate.org.uk

Abstract. The fields of Digital Humanities and Digital Preservation are not yet enjoying the full potential of the Semantic Web and relevant technologies, largely due to the highly contextualized nature of their source materials. This paper addresses the issue of representing context and use-context (i.e. context of use) of digital content, by proposing an ontology-based representation approach, which is based on the LRM, an upper-level ontology for describing dependencies between digital resources.

Keywords: Ontologies · Digital Preservation · Linked Resource Model · Context · Use-context

1 Introduction

The rapid advances in the domain of the Semantic Web and relevant technologies have yet to be leveraged in the field of Digital Humanities (DH), most probably due to the highly contextualized nature of their source materials [1]. *Digital Preservation* (*DP*) is a field in DH aimed at ensuring that digital information remains accessible and, thus, focuses on solutions that scale well along the temporal dimension.

Similarly to other fields within DH, DP does not yet enjoy the full potential of the Semantic Web, and faces a number of additional semantic challenges, like e.g. semantic and cultural ageing [2]. An array of recent research approaches attempts to address these challenges by providing solutions for dynamically discovering and invoking appropriate preservation services via Web Services [3], preserving the intelligibility/interpretability of digital objects [4], and effectively managing DP workflow risks via semantic risk management frameworks [5].

Building upon these approaches, the *PERICLES* project[1] aims to address the challenge of ensuring that digital content remains accessible in an environment that is subject to continual change. In this setting, one of the key challenges addressed by PERICLES involves the representation of the *use-context* of digital objects (DOs), which entails information related to contexts of use of the DOs. This paper proposes an ontology-based approach for representing use-context, the extraction and analysis of which relates to issues like variations of digital content interpretations and can lead to deriving meaningful correlation links among content objects and use-contexts.

2 Related Work

Early studies discriminate context modelling approaches between key-value pairs, markup, graphical, object-oriented, logic-based and ontology-based, and identify key requirements: distributed composition, partial validation, quality of information, incompleteness and ambiguity, formality and applicability [6]. Another survey adopts the six aforementioned modelling approaches, but redefines simplicity, flexibility, extensibility, genericity and expressiveness as requirements [7]. A more recent study identifies key-value pairs and markup as outdated and less expressive. Thus, it considers modelling approaches as either object-role-based, spatial, ontology-based or hybrid, while key requirements are heterogeneity, mobility, relationships, timeliness, imperfection, reasoning, usability and efficiency [8]. Indeed, an investigation of existing models in literature, reveals that the most dominant approaches are either ontology-based [9, 10, 11] or graphical [12, 13].

Regarding content and domain, most context-modelling approaches so far revolve around the topic of pervasive computing, ambient intelligence and context-aware systems, such as smart homes [14], smart meetings [15], and less often museums [16] and eLearning domains [11]. Overall, the most common concepts in context modelling tend to be *Person* and *Device* [9, 16, 17]. Examining approaches per domain, the ones in pervasive computing typically consider environmental parameters (e.g. weather, temperature, light and sound), location, user preferences, applications and services [13]. Approaches in the museum domain consider smart tour guides, but still model persistent items such as exhibits, exhibitions, artwork and media [18], as does the proposed model for DP. However, context of use is only captured in [11], which considers eLearning items used during learning design or certain activities. On the other hand, this work is aimed at DP and, therefore, has to consider rather different concepts, such as persistent content, dependencies and context of use.

3 The Linked Resource Model (LRM)

The *Linked Resource Model* (*LRM*) [19] is an upper-level ontology designed to provide a principled way for modelling evolving ecosystems, and related occurring changes. This means that, in addition to existing preservation metadata models which

[1] PERICLES FP7 project: http://www.pericles-project.eu/

aim to ensure that records remain accessible and usable over time, the LRM also aims to model how changes to the ecosystem can be captured, along with their impact. We assume that a policy governs the dynamic aspects related to changes at all times (e.g. conditions required for a change to happen and/or impact of changes). As a consequence, LRM's properties are dependent on the policy being applied and, thus, most of the defined concepts are related to what the policy expects.

At its core, the LRM defines the ecosystem by means of participating *resources* and *dependencies* among them – the `lrm` prefix refers to the LRM namespace. *Resources* represent any physical, digital, conceptual, or other kind of entity in the universe of discourse of the LRM. A resource can be either *abstract*, representing the abstract part of a resource (e.g. the idea or concept of an artwork), or *concrete*, representing the physical extension of an entity. These entities can be related through the `lrm:realizedAs` predicate, expressing, for example, that a video file is an element of the concrete realization of an abstract art piece. An abstract resource can be connected to more than one concrete resources through a container class, `lrm:AggregatedResource`.

Dependencies constitute the core concept of the LRM and describe the context under which change in one or more entities has an impact on other entities of the ecosystem. Besides the involved entities (indicated by properties `lrm:from` and `lrm:to`, which also indicate the directionality of the dependency), the description of a dependency also includes its intention and specification, as described later.

4 Representing Context and Use-context

In order to validate the models and their applicability, we have integrated the representations adopted into the domain-specific ontologies developed within PERICLES, which model resources relevant to the digital preservation of (a) digital video art (*DVA*), (b) software-based art (*SBA*), and, (c) born-digital archives (*BDA*). For the representation of digital entities, we reuse and extend several constructs from LRM, as described subsequently. More comprehensive descriptions of the ontologies can be found in [19, 20].

In the domain ontologies, the notion of `lrm:Dependency` is adopted for representing relations between digital objects and associated entities (e.g. media players, relevant software, etc.) that may further affect the functioning or display or existence of a DO. We extend the aforementioned notion into: (a) *Hardware dependencies*, which specify hardware requirements for a resource; (b) *Software dependencies* that indicate the dependency of a resource or activity on a specific software; (c) *Data dependencies*, which imply the requirement of some knowledge, data or information (e.g. passwords, configuration files, input from web services, etc.).

We represent context via associations between key classes `lrm:Agent`, `lrm:Activity` and `lrm:Resource`. More specifically, when relating an activity to a resource, the latter can be either (a) the resource that is affected by the activity, or (b) a resource that was used during the activity execution. In other words, a target resource is the one mainly handled by the activity (e.g. created, borrowed, destroyed), while used resources are those manipulated for the activity execution (e.g. equipment, software, hardware, etc.).

On the other hand, the representation of use-context capitalises on `lrm:Dependency` and its associated notions of *intention* that specifies what a dependency intends to express, and *specification* that thoroughly describes the dependency itself and its context. Thus, in order to turn dependencies into meaningful correlation links among resources and use-contexts, we have added a set of predefined intention types for representing all relevant dependency occasions. Below is a description of the proposed intention types:

- *Dependencies with a conceptual intention* are aimed at modelling the intended “meaning” of the resource (e.g. artwork) by its creator, according to the way he/she meant for the artwork to be interpreted/understood.
- *Dependencies with a functional intention* represent relations relevant to the proper, consistent and complete operation of the resource.
- *Dependencies with a compatibility intention* model compatible software or hardware components which may operate together or as substitutional components for availability, obsolescence or other reasons.

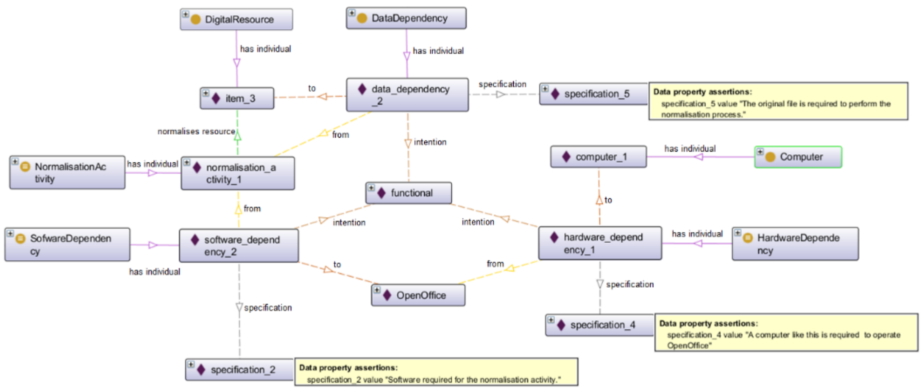

Fig. 1. All three types of dependencies combined together to attribute a complicated scenario of dependencies with functional intention.

Fig. 1 displays an instantiation example from the BDA domain that demonstrates how the representation of use-context in our models is achieved through `lrm:intention` and `lrm:specification`. The scenario represents the normalisation activity applied on a digital item, e.g. on a text file (see `item_3` in Fig. 1). Through the normalisation process, an access file is created from the initial one; the original file is in the format used by the creator, while the access format of the created file is defined by the archival policy, followed by the normalisation software used. In terms of the BDA ontology, this instantiates a software dependency of the normalisation activity on the used software. Additionally, there is a hardware dependency of the normalisation software on the hardware required in order for the software to run. The overall normalisation process also depends on the existence of the initial text file, and this information can be presented through a respective data dependency.

The intention of all three types of dependencies appearing in the figure is *functional*, meaning that all the required resources modelled in this example affect the functionality of the resources for which the dependencies were implemented (see `lrm:from` property). Also, for the sake of brevity, all hardware requirements are summarised as a computer system (*computer_1*) capable of running the software. However, the example could be easily extended to show hardware dependencies from specific components (such as CPU speed, RAM space, etc.).

5 Conclusions and Future Work

The paper proposes a novel, ontology-based representation for modelling context and *use-context* of digital resources in the DP field. At the core of the proposed representation lies the LRM, an upper ontology for modelling dependencies between DOs. Dependencies in the LRM are explicitly augmented with rich semantics, for modelling the underlying preconditions, intentions, specifications and impacts. Capitalising on these constructs, our scheme proposes a set of predefined dependency and intention types that efficiently represent the context and use-context of DOs. A sample instantiation demonstrating the introduced notions was also presented in the paper.

Our future goals are aimed at taking full advantage of LRM's capabilities, by adopting its dynamic schema (the current work considers only the static LRM part), in order to further enrich the representation of the context of use. Another aim is to investigate the implementation of mechanisms for automated use-context extraction. This could be implemented in collaboration with suitable software tools (e.g. PET [21]) for extracting information from the environment of a DO.

Acknowledgments. This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no. 601138.

References

1. Meroño-Peñuela, A. (2013). Semantic web for the humanities. In *The Semantic Web: Semantics and Big Data* (pp. 645-649). Springer Berlin Heidelberg.
2. Schlieder, C. (2010). Digital heritage: Semantic challenges of long-term preservation. *Semantic Web, 1*(1, 2), 143-147.
3. Hunter, J., & Choudhury, S. (2004, June). A semi-automated digital preservation system based on semantic web services. In *Digital Libraries, 2004. Proceedings of the 2004 Joint ACM/IEEE Conference on* (pp. 269-278). IEEE.
4. Marketakis, Y., & Tzitzikas, Y. (2009). Dependency management for digital preservation using semantic web technologies. *Int. Journal on Digital Libraries, 10*(4), 159-177.
5. Engen, V., Veres, G., Crowle, S., Bashevoy, M., Walland, P., & Hall-May, M. (2015). A Semantic Risk Management Framework for Digital Audio-Visual Media Preservation.

6. Strang, T., & Linnhoff-Popien, C. (2004, September). A context modeling survey. In *Workshop Proceedings*.
7. Baldauf, M., Dustdar, S., & Rosenberg, F. (2007). A survey on context-aware systems. *International Journal of Ad Hoc and Ubiquitous Computing*, *2*(4), 263-277.
8. Bettini, C., Brdiczka, O., Henricksen, K., Indulska, J., Nicklas, D., Ranganathan, A., & Riboni, D. (2010). A survey of context modelling and reasoning techniques. *Pervasive and Mobile Computing*, *6*(2), 161-180.
9. Zhang, D., Gu, T., & Wang, X. (2005). Enabling context-aware smart home with semantic web technologies. *Int. Journal of Human-friendly Welfare Robotic Systems*, *6*(4), 12-20.
10. Chen, H., Finin, T., & Joshi, A. (2005). The SOUPA ontology for pervasive computing. In *Ontologies for agents: Theory and experiences* (pp. 233-258). Birkhäuser Basel.
11. Jovanović, J., Gašević, D., Knight, C., & Richards, G. (2007). Ontologies for effective use of context in e-learning settings. *Educational Technology & Society*, *10*(3), 47-59.
12. Strimpakou, M., Roussaki, I., Pils, C., Angermann, M., Robertson, P., & Anagnostou, M. (2005). Context modelling and management in ambient-aware pervasive environments. In *Location-and Context-Awareness* (pp. 2-15). Springer Berlin Heidelberg.
13. Gu, T., Wang, X. H., Pung, H. K., & Zhang, D. Q. (2004, January). An ontology-based context model in intelligent environments. In *Proceedings of communication networks and distributed systems modeling and simulation conference* (Vol. 2004, pp. 270-275).
14. Ranganathan, A., McGrath, R. E., Campbell, R. H., & Mickunas, M. D. (2003, August). Ontologies in a pervasive computing environment. In *Workshop on Ontologies in Distributed Systems at IJCAI, Acapulco, Mexico*.
15. Simons, C., & Wirtz, G. (2007). Modeling context in mobile distributed systems with the UML. *Journal of Visual Languages & Computing*, *18*(4), 420-439.
16. Achilleos, A., Yang, K., & Georgalas, N. (2010). Context modelling and a context-aware framework for pervasive service creation: A model-driven approach. *Pervasive and Mobile Computing*, *6*(2), 281-296.
17. Ou, S., Georgalas, N., Azmoodeh, M., Yang, K., & Sun, X. (2006, July). A model driven integration architecture for ontology-based context modelling and context-aware application development. In *Model Driven Architecture–Foundations and Applications* (pp. 188-197). Springer Berlin Heidelberg.
18. Van den Bergh, J., & Coninx, K. (2006). Cup 2.0: High-level modeling of context-sensitive interactive applications. In *Model Driven Engineering Languages and Systems* (pp. 140-154). Springer Berlin Heidelberg.
19. Vion-Dury, J.-Y., Lagos, N., Kontopoulos, E., Riga, M., Mitzias, P., Meditskos, G., Waddington, S., Laurenson, P. and Kompatsiaris, I. (2015). Designing for Inconsistency - The Dependency-based PERICLES Approach. In T. Morzy, P. Valduriez, L. Bellatreche (Ed.), *New Trends in Databases and Inf. Systems*, 539, (pp. 458-467). Springer Berlin Heidelberg.
20. Mitzias P., Riga, M., Waddington, S., Kontopoulos, E., Meditskos, G., Laurenson, P. and Kompatsiaris, I. (2015). An Ontology Design Pattern for Digital Video. In *Proc. of the 6th Workshop on Ontology and Semantic Web Patterns (WOP 2015) co-located with the 14th Int. Semantic Web Conf. (ISWC 2015)*. Vol. 1461. Bethlehem, Pensylvania, USA.
21. Corubolo, F., Eggers, A. G., Hasan, A., Hedges, M., Waddington, S., & Ludwig, J. (2014). A pragmatic approach to significant environment information collection to support object reuse. *IPRES 2014*, 249.