Semantic Web 1 (0) 1–5
IOS Press

Editorial: Special Issue on Semantic Web for Cultural Heritage

Antonis Bikakis [a], Eero Hyvönen [b], Stéphane Jean [c], Béatrice Markhoff [d] and Alessandro Mosca [e]
[a] *Department of Information Studies, University College London, London, United Kingdom*
E-mail: a.bikakis@ucl.ac.uk
[b] *Helsinki Centre for Digital Humanities (HELDIG) and Aalto University, Finland*
E-mail: eero.hyvonen@aalto.fi
[c] *LIAS, ISAE-ENSMA and University of Poitiers, Poitiers, France*
E-mail: jean@ensma.fr
[d] *LIFAT, University of Tours, Blois, France*
E-mail: beatrice.markhoff@univ-tours.fr
[e] *KRDB Research Centre, Free University of Bozen-Bolzano, Bolzano, Italy*
E-mail: alessandro.mosca@unibz.it

Editors: Pascal Hitzler, Kansas State University, USA; Krzysztof Janowicz, University of California, Santa Barbara, USA

Abstract. Cultural Heritage and Digital Humanities have become major application fields of Linked Data and Semantic Web technologies. This editorial introduces the special issue of the Semantic Web (SWJ) journal on Semantic Web for Cultural Heritage. In total 30 submissions for the call of papers were received, of which 11 were selected for publication. The papers cover a wide spectrum of modelled topics related to language, reading and writing, narratives, historical events and cultural artefacts, while describing reusable methodologies and tools for cultural data management. This issue indicates and demonstrates the high potential of Semantic Web technologies for applications in the Cultural Heritage domain.

Keywords: Semantics, Ontologies and information modelling, Data access and exploration, Knowledge Graphs and Linked data, Cultural Heritage, Digital Humanities

1. Introduction

Cultural Heritage plays a central role to better understand previous generations and the history of where humankind comes from, and to envision where it is going to. The Web allows people to publish, explain, and debate at all scales, local, national, and worldwide. Scientific researchers, organisations, associations, and schools are looking for relevant technologies for annotating, integrating, sharing, accessing, visualising, and analysing the mine of cultural collections and, more generally, cultural data. There is also a need for taking into account profiles and preferences of end users in order to offer them highly personalised digital experiences. Several national and international research, innovation, and infrastructure programmes, such as EUROPEANA [1], DARIAH [2], PARTHENOS [3], CrossCult [4], ARCHES [5], and ARIADNEplus [6], have been launched to these directions. During 2018, which was the European Year of Cultural Heritage [7], several events and initiatives across Europe encouraged people to engage, explore and debate our rich and diverse Cultural Heritage.

When dealing with scholarly data, the "FAIR guiding principles for scientific data management and stewardship" [8] of publishing Findable, Accessible, Interoperable, and Re-usable data are a common norm. A fundamental challenge that many of the aforementioned projects deal with is how to make Cultural Heritage data, which is made available by different actors

in different cultural domains and in a multitude of different languages and formats, mutually interoperable, so that it can be searched, linked, and presented in a harmonised way across the boundaries of the datasets and data silos.

Early solutions were based on the syntactic or structural level of data, without leveraging the rich semantic structures underlying the content. During the two last decades, solutions based on the principles and technologies of the Semantic Web have been proposed to explicitly represent the semantics of data sources and make both their content and their semantics machine operable and interoperable. In parallel, knowledge representation models have matured, such as the CIDOC-CRM's ecosystem of the museum sector, interlinked with FRBR-based models in libraries, which is dedicated to the cultural heritage area including the fields of documentation, archaeology, history, architecture, etc. As more and more institutions bring their data to the Semantic Web level, the tasks of data integration, sharing, analysis, visualisation, etc. are now to be conceived in this very rich framework. At the same time, Artificial Intelligence based methods are increasingly used both in semantic content creation and in the development and support of applications for human users.

This special issue has offered to Computer Scientists, Data Scientists and Digital Humanities researchers who are involved in the development or deployment of Semantic Web solutions for Cultural Heritage the opportunity to present their realisations, the outcomes of their projects, being either publicly reusable Semantic Web tools, datasets or ontologies published in the Linked Open Data Cloud, or Semantic Web techniques, services and architectures for Cultural Heritage.

2. Special issue contributions

From the large variety of the 30 submissions coming from 16 different countries, the following papers were accepted to be included in the special issue at hand. The papers are classified in three categories, according to the main area or domain of contribution.

2.1. Language, Reading, Writing

The paper "*Ce qui est écrit et ce qui est parlé. CRMtex for modelling textual entities on the Semantic Web*" presents CRMtex, an ontology for modelling texts, with an emphasis on its more recent development. It describes the design rationale of the ontology, its classes and properties and their relations to elements of CIDOC-CRM. As demonstrated with examples, the proposed ontology is able to model not only texts as physical and linguistic entities, but also activities and procedures related to them, such as their production, transcription and decoding. Built as an extension of CIDOC-CRM, CRMtex achieves its aim of being an interoperable data model, which can be used for different types of texts and for different purposes.

The paper "*Modeling Execution Techniques of Inscriptions*" complements the previous article by focusing on the description of writing execution techniques. The limitations of the EAGLE and CRMtex models for this kind of information are explained and addressed by the authors' proposal, which can be combined with these models. Interestingly, this ontology (again extending CIDOC-CRM) complements EAGLE Vocabularies (expressed in SKOS) with a class structure.

In "*Understanding the phenomenology of reading through modelling*", the authors address another human activity, strongly related to texts or inscriptions: reading. It discusses the design process of an ontology that models the human's experience of reading, called READ-IT, carried out by an international multidisciplinary research team. This ontology is meant to semantically annotate sources of studies about reading events. As in most of the approaches presented in this special issue, the CIDOC-CRM is extensively reused in READ-IT. The resulting ontology is available in a GitHub repository.

2.2. Narratives, History, Archaeology

The paper "*Of Lions and Yakshis: Ontology-based Narrative Structure Modelling for Culturally Diverse Folktales*" describes an ontology for folk tales, which is based on Vladimir Propp's theory "Morphology of the Folktale". The aim of the ontology is to assist the analysis of folk tales by humanities researchers. The paper describes the data modelling approach, the design and the implementation of the ontology. It also presents a tool for semi-automatic extraction of information from folk tales (in textual form) and describes how the proposed ontology was applied for the analysis and comparison of African and Indian folk tales.

The notion of narratives is at the core of the paper "*Representing Narratives in Digital Libraries: The Narrative Ontology*", which presents a general approach for modelling narratives. It introduces a formal

expression of the concept of "narrative" and the resulting "Narrative Ontology" (NOnt) is the first-order logic-based counterpart of this expression. NOnt is an extension of well-known standards like CIDOC-CRM, FRBRoo and the W3C Time Ontology, and it is currently implemented with the SWRL rule language. The formalisation effort behind the development of NOnt has also given rise to the implementation of a semi-automatic software, named "Narrative Building and Visualising Tool" (NBVT), to create narratives and visualise them in several ways. The implementation of this tool is based on Semantic Web technologies and its main features are also presented in this contribution.

The paper "*WarSampo knowledge graph: Finland in the Second World War as Linked Open Data*" presents a shared knowledge graph, semantic infrastructure, and Linked Open Data service for publishing data about the World War II. The knowledge graph and data service have been used, e.g., for implementing the in-use semantic portal "*WarSampo–Finnish WW2 on the Semantic Web*" that has had hundreds of thousands of users on the Web. The system is based on representing war as a spatio-temporal sequence of events that soldiers, military units, and other actors participate in different roles. To support sustainability of the knowledge graph, a data transformation and linking pipeline has been created. The WarSampo knowledge graph, totalling approximately 14 million triples, is openly available as a service on the Linked Data Finland platform, and is part of the international LOD Cloud.

The paper "*A challenge for historical research: making data FAIR using a collaborative ontology management environment (OntoME)*" argues that the application of the FAIR data principles in the field of historical research requires the development and use of a standard ontology. It proposes adopting CIDOC-CRM as the core ontology for this domain, in combination with two other foundational ontologies, C.DnS and DOLCE. It also argues about the need of a collaborative web environment, which will enable researchers to commonly develop specifications of the core ontology for specific sub-domains or applications and align the different metadata models used by different projects. Finally it explains how the ontology management environment OntoMe can serve this purpose.

In the field of archaeology, the paper "*OntoAndalus: an ontology of Islamic artefacts for terminological purposes*" adopts the same foundational ontologies (DOLCE+DnS Ultralite (DUL)) as the previous one. It presents an ontology for Andalusian pottery artefacts, called OntoAndalus, built as a specialisation of DOLCE+DnS Ultralite (DUL) and modelled in OWL. Its development relied on the interpretation of a corpus from Portuguese and Spanish domain specific texts, English textbooks and reference works, as well as from more specialised documents from related conferences and journals articles. The paper describes the main design patterns regarding the modelling of artefact types, events and tasks, and uses the case study of *Vaso de Tavira* to exemplify how these patterns were applied to model lighting artefacts, the life cycle of pottery and the several descriptions of the artefact.

2.3. Tools for Data Management: Designing, Querying, Analysing

The ideas and approach of the "*Pattern-based design applied to cultural heritage knowledge graphs*" paper are rooted in the lessons learned, the methodologies and the modelling choices discovered during the development of ArCo, a knowledge graph consisting of a network of (EDM and CIDOC-CRM-aligned) ontologies that model the Cultural Heritage domain and a Linked Open dataset of around 172.5M triples about Italian cultural properties. The paper argues about the advantages of embracing the "eXtreme Design (XD)" in the creation of Cultural Heritage ontologies, a methodology inspired by the Extreme Programming (XP) software development approach. It provides the details behind the modelling of the ArCo ontology network, the architectural patterns in place and the characteristics of the evaluation that has been performed.

The paper "*Applying and Developing Semantic Web Technologies for Exploiting a Corpus in History of Science: the Case Study of the Henri Poincaré Correspondence*" presents a semantic virtual research environment dedicated to Henri Poincaré's letters digital corpus, i.e. letters with descriptive, scientific and mathematical content. Semantic Web technologies are used to enhance both annotation and querying of this corpus. Concerning the semantic annotation, RDFS entailment is leveraged to propose a ranked list of potential values for the RDF triples associated to specific parts of the letters. For querying, transformation rules on SPARQL queries are defined to support approximate searches on vague concepts such as "*the end of the 19th century*", which is a recurrent need in the Cultural Heritage context.

Another correspondence of a mathematician is one of the use cases of Gravsearch, a system that supports complex searches in virtual graphs, introduced

in "*Gravsearch: transforming SPARQL to query humanities data*". Gravsearch is a SPARQL query rewriting system which aims at supporting both the developers and the users with the introduction of an abstraction layer on top of the existing triplestore implementations. Gravsearch has been developed as part of the Knora (Knowledge Organization, Representation, and Annotation) API, an application by the "Data and Service Center for the Humanities" (DaSCH) whose main focus is on the preservation and promotion of digital data in the Humanities through dedicated data management methods, data storage solutions and data access platforms.

3. Summary and future directions

The variety and large number of papers submitted to this special issue suggest that Linked Data and Semantic Web technologies are becoming increasingly important in creating, publishing, and analysing Cultural Heritage data in Digital Humanities. As more and more data is becoming available in harmonised interoperable datasets, more and more intelligent applications for searching, exploring, and analysing semantically structured cultural data are also emerging. The contributions in this special issue witness this development.

Another interesting observation is that all papers in this special issue, as well as the vast majority of the papers that were not accepted, rely on existing standard ontologies and data models for the representation and interchange of cultural data, such as CIDOC-CRM, FRBR and EDM. The usage of upper level ontologies (e.g., DOLCE), ontology design patterns (e.g., c.DnS), and domain-independent standards (e.g., SKOS and DCMI), is also well testified. In this way they adopt one of the re-usability principles of FAIR data, which recommends that "(meta)data meet domain-relevant community standards". But also, very importantly, they ensure the soundness and quality of their approaches, and exploit the methods and tools that have been developed for such models.

Although several significant steps have already been made, there are still several hindrances for fulfilling the potential of Semantic Web technologies in Cultural Heritage. One of them is that, although there are now well-established ontologies for the Cultural Heritage domain and most of the related fields, there are only few tools that humanities scholars, museum practitioners and other people working in this domain, can easily use to model, manage, analyse and interlink cultural data. Some of the papers presented in this special issue attempt to address this problem and present some very promising results. There is still, though, much more work to be done, and an important lesson learned from the introduced projects, is that the involvement of the people who work in this domain both in the design and the evaluation of such tools is essential for ensuring that they will fulfill their design purposes.

As it is also obvious from the papers presented in this special issue, most of the current research in this field is still led by universities or research institutes, some of them in collaboration with large cultural heritage organisations. Smaller organisations are left behind. According to two studies on the adoption of open data practices [9, 10] (which are closely related with the FAIR and Linked Data ones), the main challenges that they face are the extra time, efforts and costs required for the digitisation of their collections, their proper documentation and rights clearance; the lack of metadata for their collections; and the lack of relevant skills among their staff. Of course, the digitisation of their collections is not a problem that Semantic Web technologies can solve. However, Semantic Web researchers can help alleviate some of these challenges by developing and providing well-documented tools and detailed guidelines on how the FAIR or Open or Linked Data principles can be applied, but also platforms that will enhance the communication, information exchange, collaboration and networking between the cultural institutions.

Acknowledgements

We would like to thank the Editors-in-Chief Pascal Hitzler and Krzysztof Janowicz for their support in compiling this special issue. The following list includes the names of the reviewers (in alphabetical order) who participated in the review process of the manuscripts submitted to the special issue. Their effort and excellent work have been fundamental for the publication of the high quality scientific contributions composing the present Semantic Web Journal volume.

Reviewers: Trond Aalberg, Valentina Bartalesi, Sotiris Batsakis, Stefano Borgo, Carmen Brando, George Bruseker, Giannis Chantas, Brice Chardin, Benjamin Cogrel, Roberto Confalonieri, Olivier Corby, Gayo Diallo, Oyvind Eide, Achille Felicetti, Roberta Ferrario, Géraud Fokou, Pietro Galliani, Pawel Garbacz, Andrea Giovanni Nuzzolese, Günther

Görz, Anais Guillem, Allel Hadjali, Fayçal Hamdi, Mikko Koho, Kalliopi Kontiza, Efstratios Kontopoulos, Nikolaos Lagos, Pietro Liuzzo, Martin Lopez Nores, Claudio Masolo, Francesco Massucci, Carlo Meghini, Albert Meroño, Amedeo Napoli, Yannick Naudet, Franco Niccolucci, Laura Pandolfo, Rafael Peñaloza, Tiago Prince Sales, Cedric Pruski, Martin Rezk, Catherine Roussey, Guillem Rull, Emilio M. Sanfilippo, Daniele F. Santamaria, Christoph Schlieder, Michalis Sfakakis, Sofia Stamou, Ranka Stankovic, Maria Theodoridou, Konstantin Todorov, Jouni Tuominen, Yannis Tzitzikas, Genoveva Vargas-Solar, Fabio Vitali, Andreas Vlachidis, Holly Wright.

References

[1] The EUROPEANA web portal. https://www.europeana.eu/.

[2] The Digital Research Infrastructure for the Arts and Humanities (DARIAH). https://www.dariah.eu/.

[3] Pooling Activities, Resources and Tools for Heritage E-research Networking, Optimization and Synergies (PARTHENOS). https://www.parthenos-project.eu/.

[4] The CrossCult web portal. https://www.crosscult.eu/.

[5] The web portal of the ARCHES Platform. https://www.archesproject.org/.

[6] The ARIADNEplus web portal. https://ariadne-infrastructure.eu/.

[7] The European Year of Cultural Heritage 2018. https://europa.eu/cultural-heritage/european-year-cultural-heritage_en.html.

[8] The FAIR Guiding Principles for scientific data management and stewardship. https://www.go-fair.org/fair-principles/.

[9] E. Foundation, Transforming the world with culture: Next steps on increasing the use of digital cultural heritage in research, education, tourism and the creative industries, Technical Report, Europeana Foundation, 2015.

[10] B. Estermann, Diffusion of Open Data and Crowdsourcing among Heritage Institutions: Results of a Pilot Survey in Switzerland, *Journal of theoretical and applied electronic commerce research* **9** (2014), 15–31. doi:10.4067/S0718-18762014000300003.