Editorial: Special Issue on Semantic Web and Ontology Design for Cultural Heritage

ANTONIS BIKAKIS, University College London, U.K.
ROBERTA FERRARIO, ISTC-CNR, Trento, Italy
STÉPHANE JEAN, University of Poitiers – ENSMA, France
BÉATRICE MARKHOFF, University of Tours, France
ALESSANDRO MOSCA, Free University of Bozen-Bolzano, Italy
MARIANNA NICOLOSI ASMUNDO, University of Catania, Italy

CCS Concepts: • **Information systems → Semantic web description languages**; **Data management systems**; • **Applied computing → Arts and humanities**; • **Computing methodologies → Ontology engineering;**

Additional Key Words and Phrases: Ontologies, Knowledge Graphs, Cultural Heritage, Digitial Humanities

ACM Reference format:
Antonis Bikakis, Roberta Ferrario, Stéphane Jean, Béatrice Markhoff, Alessandro Mosca, and Marianna Nicolosi Asmundo. 2023. Editorial: Special Issue on Semantic Web and Ontology Design for Cultural Heritage. *ACM J. Comput. Cult. Herit.* 16, 3, Article 54 (October 2023), 5 pages.
https://doi.org/10.1145/3626254

1 INTRODUCTION

After an ex-ante impact assessment of a **European Collaborative Cloud for Cultural Heritage (ECCCH)** [2], a Horizon Europe call[1] is running for Culture, Creativity, and Inclusive Society that asks for a cloud enhancing "the ability of cultural heritage actors to interact across disciplinary, institutional, sectorial and political boundaries and cooperate effectively in advancing research on cultural heritage and in developing innovative solutions for the discovery, recovery, conservation, digitalisation and valorisation of digital, digitised and digitisable cultural heritage objects." Interestingly, the ex-ante report and this call specify that such a cloud should "enable semantic representation of multiple data types (various incarnations of 2D and 3D media, video, text), stored in federated repositories according to FAIR principles, and encoding data provenance." Other European initiatives

[1] https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/opportunities/topic-details/horizon-cl2-2023-heritage-eccch-01-01

Authors' addresses: A. Bikakis, Department of Information Studies, University College London, Gower Street, WC1E 6BT, London, UK; e-mail: a.bikakis@ucl.ac.uk; R. Ferrario, Institute of Cognitive Sciences and Technologies - National Research Council (CNR), via alla Cascata 56C, 38123 Trento, Italy; e-mail: roberta.ferrario@cnr.it; S. Jean, University of Poitiers, ISAE-ENSMA Poitiers, LIAS, Téléport 2 -1 avenue Clément Ader BP 40109 86961 Chasseneuil France; e-mail: jean@ensma.fr; B. Markhoff, UMR 7324 CITERES-LAT, University of Tours, 40 rue James Watt, 37200 Tours, France; e-mail: beatrice.markhoff@univ-tours.fr; A. Mosca, KRDB Research Centre for Knowledge and Data, Faculty of Engineering, Free University of Bozen-Bolzano, Piazza Domenicani, 3, I-39100 Bozen-Bolzano BZ, Italy; e-mail: mosca@inf.unibz.it; M. N. Asmundo, Department of Mathematics and Computer Science, University of Catania, Viale A. Doria 6, I 95125, Building 5 (DMI), Catania, Italy; e-mail: marianna.nicolosiasmundo@unict.it.

© 2023 Copyright held by the owner/author(s).
1556-4673/2023/10-ART54 $15.00
https://doi.org/10.1145/3626254

multiply efforts towards Web infrastructures based on, or including more and more formally defined knowledge: Europeana has started building the common European Data Space for Cultural Heritage (the Data Space),[2] FAIRcore4EOSC[3] and FAIR-IMPACT[4] are developing coordinated solutions for greater and more harmonised use of semantic artefacts, leading to semantic interoperability within and between disciplines, including the numerous and varied fields of Cultural Heritage. At the system architecture level as well, SOLID[5] is based on the Semantic Web standards.

This growing interest in scaling up robust semantic solutions is at the service of "Humanist Knowledge" building, in line with Reference [4], i.e., where knowledge representation and the Semantic Web provide the elements of socio-technical infrastructures in which humanists and professionals of Cultural Heritage can pursue specialist work but with bridges between different research areas, professional sectors, and disciplines. On the Ontology Design side, while the CIDOC CRM ecosystem is now well recognized as the heart of interoperability in the field of cultural heritage, there is no slowing down when it comes to the formal and necessarily careful definition of the very diverse concepts used in this multi-disciplinary field. This special issue, as well as the associated series of **Semantic Web and Ontology Design for Cultural Heritage (SWODCH)** workshops,[6] show that the multidisciplinary scientific community working towards this aim is dynamic and productive.

2 SPECIAL ISSUE CONTRIBUTIONS

Twenty-four submissions were reviewed by a large pool of specialists, acknowledged and thanked at the end of this Editorial. Ten of these submissions were accepted and published in this Special Issue after two or three rounds of reviews and revisions. These can be classified in three categories according to their main area or domain of contribution: the quest for the best possible ontological model for various domains and applications (five articles), the development of applications based on ontologies and Semantic Web technologies (two articles), and the design of models and tools dedicated to building, or to analysing knowledge graphs, i.e., Linked Data (three articles).

2.1 Ontology Design/Ontology-based Data Modelling

The article "*Using Semantic Web to Create and Explore an Index of Toponyms Cited in Medieval Latin Geographical Works*" presents a part of the research conducted in the **IMAGO—*Index Medii Aevi Geographiae Operum*—** project, which aims to offer a comprehensive overview of Medieval and Renaissance Latin geographical literature through the implementation of Semantic Web technologies and the LOD paradigm. To achieve this goal, an ontology named IMAGO has been meticulously designed on top of CIDOC CRM and its extension FRBRoo. The primary focus of this article is to present the authors' formal representation of toponym knowledge within this ontology. The study includes a captivating case analysis of the Latin works of the Italian poet Dante Alighieri, wherein the knowledge of toponyms cited by his writings is presented in a highly accessible and user-friendly manner.

The article "*Experiential Observations: An Ontology Pattern-based Study on Capturing the Potential Content within Evidences of Experiences*" presents an ontology pattern called **E&O (Experience & Observation)**, which allows representing activities, experiences of such activities, observations, and traces (sources that document the evidence for observations) as distinct but related entities. The study is aimed at promoting interoperability and is both formally evaluated and assessed through case studies from different domains.

[2] https://pro.europeana.eu/page/data-space-deployment
[3] https://faircore4eosc.eu/
[4] https://fair-impact.eu/
[5] Specification defined by Tim Berners Lee for letting people store their data securely in decentralized web servers for data, called Pods: https://solidproject.org/
[6] See https://swodch2023.inf.unibz.it/ for the latest edition of SWODCH.

The article "*Between Written and Visual Communication: CIDOC CRM Ontology for Medieval and Early Modern European Graffiti*" introduces a novel ontology for the representation of historical graffiti concerning the Middle Age and Early Modern European periods. Historical graffiti are seen as a complementary source for understanding different aspects of past societies. However, a commonly shared, well-assessed definition of graffiti is still missing. This work is relevant and interesting because it extends the well-known CIDOC CRM standard with classes and properties contributing to a reliable ontological definition of graffiti.

Another extension of CIDOC CRM, for the domain of maritime history, is presented in *"The SeaLiT Ontology – An Extension of CIDOC-CRM for the Modeling and Integration of Maritime History Information*." The ontology was developed in the context of a research project on Mediterranean maritime history (SeaLiT), and the design requirements of the ontology were identified with the help of maritime historians participating in the project. The article also presents the implementation of a knowledge graph based on the ontology and a query interface developed on the ResearchSpace platform.

The domain of artworks' interpretations (with a focus on iconographic interpretations) is formally described by a new ontology, called ICON, presented in "*ICON: an Ontology for Comprehensive Artistic Interpretations*." The considered topic is relevant and currently attracts the interest of the DH community because of researchers' need to make explicit their claims about art. According to Panofsky's three-tier framework, in ICON, an interpretation of an artwork can be described at three different levels: the pre-iconographical, the iconographical, and the iconological level. ICON is designed by applying two knowledge engineering methodologies: SAMOD and **eXtreme Design (XD)**. Moreover, it has be defined to be compliant with well-known ontologies such as ArCo, CIDOC-CRM, VIR, DOLCE, and Simulation Ontology.

2.2 Applications of Semantic Web Technologies

The article "*The Facets of Intangible Heritage in Southern Chinese Martial Arts: Applying a Knowledge-driven Cultural Contact Detection Approach*" presents an inference-based computational model to study cultural contact in Southern Chinese martial arts through both tangible and intangible evidences (weapons, but also symbols). The full cycle of the model is addressed, from the organization of the evidence by constructing a domain ontology of martial arts' sources, to the generation of a group of datasets structured in a knowledge graph, to the generation of new knowledge by combining the domain ontology with a transversal cultural contact ontology.

A very inspiring example of Semantic Web applications and, more precisely, of Wikidata technologies, is presented in "*Wikibase Model for Premodern Manuscript Metadata Harmonization, Linked Data Integration, and Discovery*." The article presents a data model and a workflow for building an online Wikibase-based catalog that lists pre-modern manuscripts from around the world, from metadata of collections of over 35 institutional libraries, museums, and other cultural stakeholders of the USA. It illustrates the practical keys for implementing a large-scale system for a national platform to access manuscripts preserved in many different institutions. Its content clearly covers the different levels of issues, from the requirements to the technical workflow that allows integrating the data provided by the institutions, enriching it with LOD authority appellations. It also shows how to design the common model that allows aggregating data, i.e., the lowest common denominator of many manuscript descriptions, on which a common catalog can be built that efficiently redirects to the different institution web portals.

2.3 Models and Tools for Building or Analysing Knowledge Graphs, or Linked Data

It is of outmost importance for practitioners using Cultural Heritage Linked Open Data to *understand* what a knowledge graph contains. Knowledge graphs may come with an ontology but most of them actually use several ontologies and, more precisely, several parts of ontologies. Moreover, with knowledge graphs that correctly use the CIDOC CRM, knowing the schema is not enough to understand anything, because several terminologies are used at the data level to complement the description of what the data means. For instance, knowing that the graph contains *crm:E22 Human-made Objet* instances is far less useful than knowing that these instances are of type *nomisma:coin*. The system presented in the article *"TTProfiler: Types and Terms Profile Building for Online Cultural Heritage Knowledge Graphs"* generates a profile by querying the SPARQL endpoint of an online knowledge graph, and this profile allows users to visualize and navigate among the classes and properties used and to visualize the terms that characterize entities in this graph.

The article *"Comparing Heuristic Rules and Masked Language Models for Entity Alignment in the Literature Domain"* focuses on the problem of entity alignment for the purposes of cross-linking within the Linked Data cloud. It particularly explores two methods, one based on a set of domain-specific heuristic rules and a deep-learning one based on **masked language models (MLMs)**. It evaluates the two methods on four heterogeneous datasets containing bibliographic records from the Quebec literature modelled according to the IFLA LRM ontology. The main conclusions of this study are that both methods achieve excellent performance, but the MLM-based approach can more easily be generalised and has looser requirements with respect to data cleaning and data structure.

Archaeological excavations are non-linear, complex, and irreversible processes. The destructive excavation kind of activity that characterises the archaeological excavation makes detailed and accurate documentation an absolute necessity. In fact, a number of heterogeneous interconnected data is dynamically produced, during and after an excavation, that includes spatial data, texts, recording sheets, 3D models, external repositories, reports, scientific papers, content for exhibitions, and other mediation applications, among others. In *"New AIR for the Archaeological Process? The Use of 3D Web Semantic for Publishing Archaeological Reports,"* the authors introduce the web platform **Archaeological Interactive Report (AIR)**, a system designed for recording archaeological investigations in real-time, an online archive that incorporates the complete dataset of the investigations, and a multimedia visualization system providing a 3D environment for data analysis and assemblages for testing interpretation hypotheses and publishing dynamic editorialization outputs. AIR was designed to specifically address the fieldwork archaeologists need for documentation, archival and online publication of investigations data through the development of a flexible and user-friendly tool. Here, semantic technologies guarantee interoperability among data coming from different excavation campaigns, and facilitate the integrated management of distinct data sources, the users' access to them and their FAIR-compliant publication on the Web.

3 SUMMARY AND FUTURE DIRECTIONS

The articles in this special issue demonstrate the rise of concrete applications of ontologies, Semantic Web technologies and Linked Data principles for CH, together with better tools to support the development of such applications. As already noted in a previous special issue [1], ontology design and data modelling in this domain are guided by the re-usability principles of FAIR data, which recommend that (meta)data should also be "FAIR-ified," without, however, paying less attention to soundness and quality, ensured to a large extent by the adoption of standard ontologies, such as CIDOC-CRM. More generally, the Semantic Web and Linked Data drive a great deal of research in CH as their aims (e.g., decentralising, sharing, opening, breaking data-silos, collaborating) are perfectly aligned with the particular needs of this domain.

Regarding future research in this area, one of the most compelling topics, to which we would like to turn the attention of scholars, concerns the challenges that recent developments in sub-symbolic artificial intelligence, such as machine learning and statistical applications, pose to formal knowledge modelling and, more generally, to the area of knowledge representation and reasoning. According to the recently published "Manifesto from Dagstuhl Perspectives Workshop 22282: Current and Future Challenges in Knowledge Representation and Reasoning" [3], among these challenges are undoubtedly those related to (i) the design of intelligent systems based on commonsense reasoning, that is, "the ability to make effective use of ordinary, everyday, experiential knowledge in achieving ordinary, everyday, practical goals," whose impact spreads from natural language understanding, to constructing visual interpretations, planning, and interacting with the real world; (ii) the development of innovative and reliable techniques and tools to support knowledge acquisition and management, "that is, for the formulation of suitable domain knowledge by knowledge engineers and domain experts"; (iii) the design of hybrid systems in which machine learning techniques (e.g., deep neural networks, Generative and Large Language Models) are integrated with the symbolic representation of knowledge and associated reasoning techniques; (iv) the possibility of giving human-intelligible explanations of predictions or recommendations made by models developed using machine learning. To reap the benefits of both formal knowledge representation and statistical machine learning, DH scholars should not ignore such challenges and the research questions they imply. We expect that the maturity gained in this field, which we directly witnessed in the SWODCH workshops and the related follow-up initiatives, will lead to original and effective solutions, tailored to the specific needs of CH applications, but also transferable to other domains with similar characteristics.

ACKNOWLEDGMENTS

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 JOCCH volume.

Trond Aalberg, Alessandro Adamou, Alessio Antonini, Salvatore Arcidiacono, Luigi Asprino, Valentina Bartalesi, Dimitris Bilidas, George Bruseker, Francesco Corcoglioniti, Marilena Daquino, Oyvind Eide, Stefano Faralli, Catherine Faron-Zucker, Achille Felicetti, Géraud Fokou, Richard Gartner, Manolis Georgatsoulis, Xavier Granier, Anais Guillem, Anas Fahad Khan, Mikko Koho, Konstantinos Kotis, Davide Lanti, Sylvain Laubé, Cristiano Longo, Marinucci Ludovica, Ludovica Marinucci, Yannis Marketakis, Claudio Masolo, Daniele Metilli, Andrea Giovanni Nuzzolese, Laura Pandolfo, Maria Angela Pellegrino, Lighton Phiri, Davide Picca, Emilio M. Sanfilippo, Daniele F. Santamaria, Michalis Sfakakis, Daria Spampinato, Sofia Stamou, Maria Rosaria Stufano Melone, Maria Theodoridou, Konstantin Todorov, Francesca Tomasi, Cássia Trojahn, Christos Tryfonopoulos, Valentina Vassallo, Costas Vassilakis, Fabio Vitali, Andreas Vlachidis, Dan Vodislav.

REFERENCES

[1] Antonis Bikakis, Eero Hyvonen, Stéphane Jean, Béatrice Markhoff, and Alessandro Mosca. 2021. Editorial: Special issue on semantic web for cultural heritage. *Semant. Web* 12, 2 (2021), 163–167.

[2] Pere Brunet, Livio De Luca, Eero Hyvönen, Adeline Joffres, Peter Plassmeyer, Martijn Pronk, Roberto Scopigno, and Gábor Sonkoly. 2022. *Report on a European Collaborative Cloud for Cultural Heritage. Ex-Ante Impact Assessment.* Technical Report. Directorate-General for Research and Innovation (European Commission). DOI: https://doi.org/10.2777/64014

[3] James P. Delgrande, Birte Glimm, Thomas Meyer, Miroslaw Truszczynski, and Frank Wolter. 2023. Current and future challenges in knowledge representation and reasoning. *arXiv preprint arXiv:2308.04161* (2023).

[4] Dominic Oldman, Martin Doerr, and Stefan Gradmann. 2015. Zen and the art of linked data: New strategies for a semantic web of humanist knowledge. *New Compan. Digit. Human* (2015), 251–273.