

Proceedings of the First International Workshop of Semantic Digital Humanities co-located with the Extended Semantic Web Conference 2024

Oleksandra Bruns^{1,2,*}, Andrea Poltronieri³, Lise Stork⁴ and Tabea Tietz^{1,2}

¹FIZ Karlsruhe – Leibniz Institute for Information Infrastructure, Hermann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen, Germany

²Karlsruhe Institute of Technology (AIFB), Kaiserstr. 89, 76133 Karlsruhe, Germany

³Department of Computer Science and Engineering, University of Bologna, Mura Anteo Zamboni 7, 40126 Bologna, Italy

⁴Informatics Institute, University of Amsterdam, Science Park 900, 1098 XH Amsterdam, The Netherlands

Abstract

Exploration, analysis, and preservation of the rich cultural and historical tapestry of the world are essential for our understanding of humanity's past and shaping our future. In recent years, there has been increased interest in the creation and application of Ontologies, Knowledge Graphs, and other Semantic Web Technologies within cultural heritage (CH) and digital humanities (DH). However, to date, the distinct areas of expertise, methodologies and traditions across the fields have led to a noticeable gap between tech solutions and humanities' needs. The aim of the International Workshop of Semantic Digital Humanities (SemDH) was to bridge this division and encourage closer collaboration and networking across diverse fields.

Keywords

Digital Humanities, Knowledge Graphs, Semantic Representation, Large Language Models, Cultural Heritage, Ontologies, FAIR

1. Organising Committee

- Oleksandra Bruns, FIZ-Karlsruhe, Karlsruhe Institute of Technology
- Andrea Poltronieri, University of Bologna
- Lise Stork, University of Amsterdam
- Tabea Tietz, FIZ-Karlsruhe, Karlsruhe Institute of Technology

2. Steering Committee

- Harald Sack, FIZ Karlsruhe, Karlsruhe Institute of Technology

First International Workshop of Semantic Digital Humanities (SemDH'24) co-located with ESWC 2024, May 27th, 2024, Hersonissos, Greece

*Corresponding author.

✉ oleksandra.bruns@fiz-karlsruhe.de (O. Bruns); andrea.poltronieri2@unibo.it (A. Poltronieri); l.stork@uva.nl (L. Stork); tabea.tietz@fiz-karlsruhe.de (T. Tietz)

🌐 <https://www.fiz-karlsruhe.de/en/forschung/lebenslauf-und-publicationen-oleksandra-bruns> (O. Bruns); lifestork.github.io (L. Stork)

🆔 0000-0002-8501-6700 (O. Bruns); 0000-0002-2146-4803 (L. Stork)

© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

- Torsten Schrade, Academy of Sciences and Literature Mainz
- Stefan Schlobach, Vrije Universiteit Amsterdam
- Valentina Presutti, University of Bologna
- Catherine Faron, Université Côte d'Azur
- Eero Hyvönen, Aalto University
- Francesca Tomasi, University of Bologna
- Mareike König, German Historical Institute Paris
- Laura Hollink, Centrum Wiskunde Informatica

3. Program Committee

- Valentina Carriero, Cefriel
- Stefano De Giorgis, University of Bologna
- Daniil Dobriy, Vienna University of Economics and Business
- Catherine Faron, Université Côte d'Azur
- Ariana Graciotti, University of Bologna
- Ivan Heibi, University of Bologna
- Eero Hyvönen, Aalto University and University of Helsinki
- Nitisha Jain, King's College London
- Mareike König, German Historical Institute Paris
- Nicolas Lazzari, University of Bologna
- Delfina Sol Martinez Pandiani, Centrum Wiskunde & Informatica
- Andrei Nesterov, Centrum Wiskunde, Informatica
- Harald Sack, FIZ Karlsruhe, Karlsruhe Institute of Technology
- Stefan Schlobach, Vrije Universiteit Amsterdam
- Torsten Schrade, Academy of Sciences and Literature Mainz
- Sarah Shoilee, Vrije Universiteit Amsterdam
- Francesca Tomasi, University of Bologna
- Mary Ann Tan, FIZ Karlsruhe, Karlsruhe Institute of Technology
- Mahsa Vafaie, FIZ Karlsruhe, Karlsruhe Institute of Technology
- Shenghui Wang, University of Twente
- Andreas Weber, University of Twente
- Viktor de Boer, Vrije Universiteit Amsterdam
- Jacco van Ossenbruggen, Vrije Universiteit Amsterdam

4. Preface

The International Workshop of Semantic Digital Humanities (SemDH'24)¹ invited members of the Semantic Web, CH, and DH communities that were actively involved in the development of

¹<https://semhdh.github.io/>

systematic approaches and advanced technologies for handling heterogeneous, diverse, and challenging humanities data. It provided a platform to researchers to share and discuss their challenges and showcase research findings. Moreover, it encouraged closer collaboration and networking across these diverse fields.

The proposed topics included, but were not limited to, constructing and utilizing knowledge graphs for the humanities, extracting and representing knowledge from historical texts, data linking across disciplines, enriching the semantics of historical records and biographies, analyzing social networks, and adopting, extending, and evaluating ontologies for specific domains. The participants were also encouraged to submit works using a combination of Semantic Technologies with more recent technologies, such as Large Language Models (LLMs), to tackle the diverse challenges associated with working with historical data.

There were 16 papers submitted for peer-review to this workshop. Out of these, 12 papers were accepted for this volume, 6 as regular papers and 6 as short papers. The range of accepted papers demonstrated a focus not only on technological endeavors, but also on the analysis of CH and DH data landscapes from diverse perspectives. Full research and resource papers were presented in two full paper presentation sessions, while work-in-progress contributions were discussed in a dedicated poster session. Additionally, the workshop featured a keynote by Marieke van Erp titled "Knowing What You Don't Know - On Gaps, Quality Issues, and Other Challenges in DH Use Cases." Following the keynote was a panel discussion with Harald Sack, Marieke van Erp, Enrico Daga, Lise Stork, Victor de Boer, and Eljas Oksanen.

The workshop's success was demonstrated not only by the high number of submissions but also by the level of engagement from participants, the quality of discussions, and the breadth of topics covered, showing the importance of the workshop in fostering collaboration and advancing the integration of semantic technologies within the domains of cultural heritage and digital humanities.

June, 2024

Oleksandra Bruns, Andrea Poltronieri, Lise Stork, and Tabea Tietz

5. Contents

Full Paper Session I

Bistiris Ontology: Towards a Structured Representation of Sardinian Traditional Female Costumes.

Giorgio Corona, Dario Guidotti, Laura Pandolfo and Luca Pulina

Exploring Prosopographical Information in the Virtual Record Treasury of Ireland's Knowledge Graph for Irish History.

Beyza Yaman, Lucy McKenna, Alex Randles, Lynn Kilgallon, Peter Crooks and Declan O'Sullivan

Publishing Numismatic Public Finds on the Semantic Web for Digital Humanities Research – CoinSampo Linked Open Data Service and Semantic Portal.

Heikki Rantala, Eljas Oksanen, Frida Ehrnsten and Eero Hyvönen

Full Paper Session II

Towards a semantic representation of Egyptian demonology: requirements and benchmark study.

Bruno Sartini and Rita Lucarelli

eXtreme Design for Ontological Engineering in the Digital Humanities with Viewsari, a Knowledge Graph of Giorgio Vasari's The Lives.

Sarah Rebecca Ondraszek, Grischka Petri, Ulrike Blumenthal, Lisa Dieckmann, Etienne Posthumus and Harald Sack

A Corpus of Biblical Names in the Greek New Testament to Study the Additions, Omissions, and Variations across different Manuscripts.

Christoph Werner, Zacharias Shoukry, Soham Al-Suadi and Frank Krüger

Keynote: “Knowing what you don't know - on gaps, quality issues and other challenges in DH use cases”

Dr. Marieke van Erp

Working with historical data is not for the faint of heart: data can be incomplete, warped through digitisation artefacts, or difficult to understand as language and society have changed. However, it also provides a window on the past that can help us understand today's society better. In this talk, I will give examples of DH use cases that take on gaps in data, quality issues and other challenges head-first to make computational methods better suited to them.

Panel Discussion: “Bridging the Gap: Semantic Web and Digital Humanities”

Prof. Dr. Harald Sack, Dr. Marieke van Erp, Dr. Enrico Daga, Dr. Lise Stork, Dr. Victor de Boer, and Dr. Eljas Oksanen.

The panel discussion dived into the intersections of Semantic Web and Digital Humanities, exploring challenges, perspectives, and strategies for fostering collaboration between these fields. Experts from both domains shared insights on collaborative research projects, digital methods, and interdisciplinary education. Key themes included the need for responsible AI, user-centric design, and the importance of aligning technological solutions with the needs of humanities researchers. The panel emphasized the necessity of mutual learning, interdisciplinary education, and ongoing dialogue to bridge the gap between Semantic Web and Digital Humanities, aiming to enhance the integration of semantic technologies within humanities research.

Poster Session for Short and Position Papers

PaleOrdia: Semantically Describing (Cuneiform) Paleography using Paleographic Linked Open Data.

Timo Homburg

Towards LLM-based Semantic Analysis of Historical Legal Documents.

Tania Litaina, Andreas Soularidis, Georgios Bouchouras, Konstantinos Kotis and Evangelia Kavakli

Sustainable Semantics for Sustainable Research Data.

Steffen Henniecke, Pascal Belouin, Hassan El Hajj, Matthew Fielding, Robert Casties and Kim Pham

Non-Canonical Acts and their Topical Distribution.

Christian Vrangbæk, Eva Vrangbæk, Márton Kardos, Kristoffer Nielbo and Jacob Mortensen

Digitalisation Workflows in the Age of Transformer Models: A Case Study in Digital Cultural Heritage.

Mahsa Vafaie, Mary Ann Tan and Harald Sack

FAIR Paper: Applying FAIR to Academic Publishing.

Wouter Beek, Rick Maurits and Auke Rijpma

6. Summary of the Panel “Bridging the Gap: Semantic Web and Digital Humanities”

The panel brought together six distinguished experts, Harald Sack, Marieke van Erp, Enrico Daga, Lise Stork, Victor de Boer, and Eljas Oksanen. With these representatives from both, the Semantic Web and (Digital) Humanities fields the goal was to discuss the challenges, gaps, perspectives, and future steps for a closer collaboration between these domains. In a two-minute statement, each panelist provided an outline of their position on the intersection of Semantic Web and Digital Humanities, sharing their experiences, and identifying the main challenges they acquired. Afterwards, the panelists were directed with concrete questions, e.g. their views on the effectiveness of Semantic Web solutions in addressing the needs of (Digital) Humanities; on main differences between the fields, and proposing strategies to bridge the gap. Additionally, the audience was actively encouraged to engage with the panelists by posing their own questions, which ranged from the reasons on possible reluctance of humanists to adopt Semantic Web technologies, e.g. Knowledge Graphs, to the difficulty of teaching SPARQL to humanists, and possible strategies for dealing with data quality and incompleteness in Digital Humanities research. The results of the discussion can be summarized in the following key challenges and the proposals on addressing them:

- **Differences in Methodologies.** The panelists emphasized the critical evaluation required when using digital methods in the humanities, acknowledging the imperfections and complexities of historical data. **Possible Solution:** A more user-centered approach, as well as the adoption of responsible AI and fostering mutual learning between fields.
- **Challenges in Digital Methods and Data Ownership.** On example of historical videos, issues such as intellectual property rights and ownership of data post-project completion

were mentioned. **Possible Solution:** Clear guidelines on data ownership and rights management in digital projects were highlighted.

- **Research Bubbles.** According to the discussion, there are growing concerns about whether fields such as Digital Humanities and the Semantic Web are effectively addressing the problems of the Humanities. The focus of Digital Humanities has shifted towards self-fulfillment rather than directly meeting the specific needs of the field. Nonetheless, they provide important tools that foster novel approaches to evidence interpretation, and humanists require knowledge of how to use these tools. Similarly, concerns have been raised about Semantic Web methods in Humanities, particularly regarding whether ontologies, often simplified representations, truly address the needs of humanities researchers. To address this, computer scientists require to possess a fundamental understanding of humanities principles and perspectives. **Possible Solution:** Interdisciplinary education and mutual learning, the development of tools that align with the needs of humanists.
- **Differences in Research Paces.** The differing research paces between Semantic Web and Humanities projects and the importance of aligning research questions with project goals were mentioned. **Possible Solution:** Closer collaboration and communication to synchronize research timelines and objectives.
- **Training and Mutual Learning.** The solution to many of the highlighted challenges lies in training humanists in computer science and fostering interdisciplinary learning. There is an ongoing debate about whether humanists should be trained in technical skills like SPARQL queries and ontology development, and how to effectively implement this training. Additionally, it was pointed out that computer scientists should also learn more about history and humanities methods. The core challenge is how to effectively provide this kind of training and develop a curriculum that bridges the gap between the fields while motivating experts to engage in mutual learning. **Possible solution:** To effectively train humanists in technical approaches, it is crucial to focus on familiarizing them with digital methods and tools, rather than requiring coding skills. In countries like Finland, digital methods are already integrated into humanists' curriculum, and many Digital Humanities students are established historians. For more experienced researchers, workshops, short courses, and clear communication emphasizing practical applications of digital tools are key. Fostering interdisciplinary collaboration through joint projects enables mutual learning between humanists and computer scientists. Developing a curriculum that showcases the practical benefits of digital methods can motivate humanists to embrace these new approaches.
- **Data Quality and Provenance:** Understanding data quality and provenance poses a significant challenge in Digital Humanities research. Without proper mechanisms in place, ensuring the accuracy and reliability of data becomes difficult. **Possible Solution:** Utilizing machine learning techniques can aid in identifying gaps in data, helping to enhance data quality. Additionally, employing SPARQL queries with proper provenance is crucial for maintaining the accuracy and reliability of data sources.

Overall, the panel underscored the need for deeper interdisciplinary collaboration, mutual education, and the development of tools and methodologies that truly meet the needs of both Semantic Web and Digital Humanities researchers.

Acknowledgments

We want to thank the contributors and the steering and program committee members for their work. This work has been supported the by Deutsche Forschungsgemeinschaft (DFG), project NFDI4Culture (project number 441958017), and EU's Horizon Europe research and innovation programme, the MUHAI project (grant agreement no. 951846).