Proceedings of the Second International Workshop of Semantic Digital Humanities co-located with the Extended Semantic Web Conference 2025

Oleksandra Bruns[1,2], Arianna Graciotti[3], Bruno Sartini[4] and Tabea Tietz[1,2]

[1]FIZ Karlsruhe – Leibniz Institute for Information Infrastructure, Hermann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen, Germany
[2]Karlsruhe Institute of Technology (AIFB), Kaiserstr. 89, 76133 Karlsruhe, Germany
[3]LILEC, University of Bologna, 40126 Bologna, Italy
[4]Ludwig-Maximilians-Universität München, Germany

Abstract

Investigating, interpreting, and safeguarding the world's cultural and historical assets are crucial for comprehending humanity's past and future. Recently, there has been a surge of interest in applying Ontologies, Knowledge Graphs, and Semantic Web Technologies to Cultural Heritage (CH) and Digital Humanities (DH). Nonetheless, varying areas of expertise and traditions have led to a disconnect between technological solutions and the needs of the humanities. The International Workshop of Semantic Digital Humanities (SemDH) aims to close this divide and promote cooperation among the Semantic Web, CH, and DH communities.

Keywords

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

1. Organizing Committee

- Oleksandra Bruns, FIZ Karlsruhe and Karlsruhe Institute of Technology, Germany
- Arianna Graciotti, LILEC, University of Bologna, Italy
- Bruno Sartini, Institute for Digital Cultural Heritage Studies, Ludwig-Maximilians-Universität München, Germany
- Tabea Tietz, FIZ Karlsruhe and Karlsruhe Institute of Technology, Germany

2. Steering Committee

- Enrico Daga, Knowledge Media Institute (United Kingdom)
- Catherine Faron, Université Côte d'Azur (France)
- Laura Hollink, Centrum Wiskunde Informatica (Netherlands)
- Eero Hyvonen, Aalto University (Finland)
- Mareike König, German Historical Institute Paris (France)
- Valentina Presutti, University of Bologna (Italy)
- Harald Sack, FIZ Karlsruhe and Karlsruhe Institute of Technology (Germany)

SemDH 2025: Second International Workshop of Semantic Digital Humanities. Co-located with ESWC 2025, June 02, 2025, Portoroz, Slovenia.
oleksandra.bruns@fiz-karlsruhe.de (O. Bruns); arianna.graciotti@unibo.it (A. Graciotti); b.sartini@lmu.de (B. Sartini); tabea.tietz@fiz-karlsruhe.de (T. Tietz)
https://www.fiz-karlsruhe.de/de/bereiche/lebenslauf-und-publikationen-oleksandra-bruns (O. Bruns); https://www.unibo.it/sitoweb/arianna.graciotti/en (A. Graciotti); https://www.fiz-karlsruhe.de/en/bereiche/lebenslauf-und-publikationen-tabea-tietz (T. Tietz)
0000-0002-8501-6700 (O. Bruns); 0009-0004-7918-809X (A. Graciotti); 0000-0002-9152-4402 (B. Sartini); 0000-0002-1648-1684 (T. Tietz)

- Stefan Schlobach, Vrije Universiteit Amsterdam (Netherlands)
- Torsten Schrade, Academy of Sciences and Literature Mainz (Germany)
- Francesca Tomasi, University of Bologna (Italy)

3. Program Committee

- Sofia Baroncini, Leibniz Institut für Europäische Geschichte (IEG)
- Stefano De Giorgis, University of Bologna
- Daniil Dobriy, Vienna University of Economics and Business
- Catherine Faron, Université Côte d'Azur
- Ivan Heibi, University of Bologna
- Nitisha Jain, King's College London, London, United Kingdom
- Shufan Jiang, FIZ Karlsruhe
- Marieke Koenig, Deutsches Historisches Institut Paris
- Nicolas Lazzari, University of Bologna
- Fabio Mariani, University of Augsburg
- Delfina Sol Pandiani, University of Amsterdam
- Harald Sack, FIZ Karlsruhe and Karlsruhe Institute of Technology
- Sarah Binta Alam Shoilee, Vrije Universiteit Amsterdam
- Gunjan Singh, FIZ Karlsruhe
- Gianmarco Spinaci, Villa i tatti, Harvard University
- Mary Ann Tan, FIZ Karlsruhe
- Francesca Tomasi, Unviersity of Bologna
- Shenghui Wang, University of Twente
- Mathias Zinnen, Friedrich Alexander Universität Erlangen-Nürnberg
- Victor de Boer, Vrije Universiteit Amsterdam
- Jacco van Ossenbruggen, VU Amsterdam

4. Preface

The SemDH workshop serves as a space for bridging the gap between the Semantic Web and the Digital Humanities (DH) and Cultural Heritage (CH) communities. This year's edition of the workshop, SemDH2025, presented a variety of novel works that explored how Linked Open Data (LOD) addresses the challenges of representing, preserving, and enhancing cultural knowledge. There were 20 papers submitted for peer-review to this workshop, each one of them was reviewed by at least 3 members of the program committee. Compared to the 2024 edition (SemDH'24) [1], the number of submissions increased by 25%. After the reviews, 12 papers were accepted for this volume, 3 as full research papers, and 9 as short or position papers. The topics of the papers covered a range of research problems, including the development of CH knowledge graphs, visualization tools for historical data, and the application of semantic technologies to literary tourism, coreference resolution for ancient languages, the creation of digital humanities research portals, etc. The technical possibilities of the Semantic Web were explored, but the discussions also touched on the need for humanistic perspectives to guide the development of systems that truly serve the needs of the DH and CH communities. Thus, a central common theme mentioned across the papers was the importance of interdisciplinary collaboration between the Semantic Web, Cultural Heritage, and Digital Humanities fields.

This year, a major topic of discussion was the growing recognition of cultural bias in LOD. The papers, the keynote, and the panel discussion highlighted the need to confront biases within CH collections and historical datasets. As cultural knowledge is encoded into digital systems, outdated and offensive terminology, as well as stereotypical representations, can be perpetuated inadvertently. The keynote by

Laura Hollink addressed these issues, analyzing the presence of such bias across structured metadata, controlled vocabularies, and knowledge graphs. The panel discussion further emphasized the critical role of addressing cultural bias within LOD, discussing practical strategies for detection and mitigation. As the workshop progresses, these discussions set the stage for continued reflection and action toward creating inclusive, fair, and culturally responsible LOD systems. Finally, taking into consideration both the initial reviews, the presentations and resulting discussions, the SemDH organizing committee decided to give the best paper award to *Enhancing Provenance Research with Linked Data: A Visual Approach to Knowledge Discovery* by Sarah Binta Alam Shoilee et al.

June 2025 Oleksandra Bruns, Arianna Graciotti, Bruno Sartini, and Tabea Tietz

5. Contents

5.1. Session I

5.2. Session II

5.3. Session III

5.4. Keynote: Cultural Bias in Linked Open Data

Laura Hollink

Cultural heritage collections often reflect the societal values and norms prevalent at the time when objects were created, collected, cataloged, and described. As a result, they may include outdated, stereotypical, or offensive terminology relating to people and cultures. In this presentation, we examine the presence of such contentious language within cultural heritage collections. Our analysis spans multiple layers: the cultural objects themselves, the structured and unstructured metadata used to describe and interlink them, and the controlled vocabularies, thesauri, and knowledge graphs that underpin these systems. Across all levels, we identify significant forms of bias. We will present methods for detecting these biases at scale, and discuss approaches for mitigation. In conclusion, we reflect on what the linked open data community can learn from cultural heritage institutions in confronting and addressing cultural bias in large-scale datasets.

5.5. Panel Discussion: Cultural Bias

- Laura Hollink, Centrum Wiskunde & Informatica (Netherlands)
- Rossana Damiano, University of Torino (Italy)
- Torsten Schrade, Academy of Sciences and Literature Mainz (Germany)
- Harald Sack, FIZ Karlsruhe and Karlsruhe Institute of Technology (Germany)
- Mrinalini Luthra, Huygens Insitute (Netherlands)
- Amber Zijlma, Huygens Institute (Netherlands)

6. Summary of the Panel on Cultural Bias

The panel on ”Cultural Bias in LOD” brought together experts from the fields of Semantic Web and Digital Humanities, including Laura Hollink, Rossana Damiano, Torsten Schrade, Harald Sack, Mrinalini Luthra, and Amber Zijlma. The audience was also considered an integral part of the panel and was strongly encouraged to participate in the discussion by sharing experiences, asking questions, and offering insights.

At the beginning of the session, the audience was asked to participate in a set of icebreaker questions, which provided valuable context and perspective on the topics at hand. Their responses helped to define the current state of cultural bias awareness and highlighted some key areas for improvement. In the following, a set of questions and the corresponding response statistics are provided:

- *What is your background?*
  The audience (21 participants) represented a diverse mix of fields, with 13 participants from Computer Science, 3 from Digital Humanities, 3 from the Humanities and 2 from other fields.
- *During your research, did you consider any cultural bias aspects?*
  The majority of participants (12) reported that they had considered cultural bias in their previous LOD projects, while 9 did not.
- *In the requirements collection phase, did you involve any stakeholders from the Humanities or end-user target group?*
  A significant portion of the audience (16) indicated that they involved stakeholders from the Humanities or end-user target groups during the requirements collection phase, while 6 did not.
- *Did you question the cultural assumptions of the ontologies or vocabularies that you (re)used?*
  The responses were evenly split, with 11 participants stating that they questioned the cultural assumptions embedded in the ontologies they used, while 11 did not.
- *Did you plan any post-project check to see how the community has adopted your tools? Do you measure the impact of your project?*
  Only 7 participants indicated that they had planned post-project checks to measure the impact of their tools on the target community, while 15 did not.

The responses provided valuable insights into the audience's awareness and practices regarding cultural bias in LOD. Most of the participants indicated that they had considered cultural bias in their projects and had involved end users during the requirement collection phase. However, responses were more divided when it came to questioning the cultural assumptions in the ontologies and vocabularies used in their projects. Additionally, there was a clear gap in post-project evaluation, with many participants not planning checks to assess how their tools were adopted or the impact they had on the target community. These findings highlight both positive progress and areas that require more attention, particularly in terms of ongoing evaluation and critical reflection on the cultural aspects of LOD systems.
These insights provided valuable context for the core discussion that followed, as the panelists focused on how biases in data representation impact the humanities and how to ensure more inclusive, diverse, and accurate knowledge representation in LOD. The results of the discussion can be summarized in the following **key themes** and **proposals**:

- **Cultural Neutrality in LOD.** The panel explored the idea of whether LOD can ever truly be culturally neutral, concluding that LOD is inherently shaped by dominant ontologies and vocabularies. The concept of neutrality itself was questioned, as striving for neutrality may already introduce bias. **Possible Solution:** instead of attempting to eliminate bias, it should be transparently documented within the data, to ensure that users understand the assumptions and cultural perspectives embedded in LOD systems.
- **Bias in Standard Vocabularies and Knowledge Representation.** It was acknowledged that the reuse of standard vocabularies can perpetuate cultural biases. These vocabularies are often inadequate for representing diverse perspectives and may contribute to the marginalization of less- dominant knowledge systems. **Possible Solution:** a key point was made that computer scientists and data modelers must recognize their own biases and how these biases affect the data structures they create.
- **Crowdsourcing and Inclusion.** Crowdsourcing was identified as both a valuable tool and a potential source of bias. The demographics of crowdsourced contributors tend to be skewed toward privileged groups in Western contexts, leading to underrepresentation of marginalized voices. Furthermore, the financial and social barriers to participation in crowdsourcing can prevent certain communities from contributing. Ensuring inclusivity in crowdsourced data collection requires careful thought about how to involve underrepresented groups. **Possible Solution:** efforts should be made to ensure more inclusive participation in crowdsourcing, including paying contributors to encourage diverse input and considering the financial and social barriers that limit access to participation.
- **Polyvocality and Simplification in Data Representation.** The panel discussed the tension between the complexity of humanistic data and the tendency to simplify it for use in LOD systems. Humanities scholars often require nuanced, polyvocal representations of knowledge, while LOD systems tend to simplify data to fit formalized structures. This simplification can reduce the richness of historical and cultural data, potentially obscuring important perspectives. **Possible Solution:** the panel emphasized the need for more complex data representations that capture the polyvocality of humanistic knowledge. LOD systems should avoid oversimplification and instead embrace the complexity that humanities scholars require.
- **Adoption of LOD Tools by Target Communities.** A key concern was whether the tools and systems created for LOD are truly adopted by the communities they are designed for. The panel discussed the gap between the goals of computer scientists and the needs of humanities scholars, noting that LOD tools often do not align with the ways in which these scholars engage with data. **Possible Solution:** Better communication and collaboration between computer scientists and humanities scholars is necessary to ensure that LOD tools meet the needs of their intended users.
- **Data Quality and Bias in Ontologies.** It was noted that cultural bias is often embedded in ontologies and knowledge graphs. The panel discussed how this bias can impact the quality and accuracy of the data, particularly in cases where marginalized voices are underrepresented or

misrepresented. **Possible Solution:** Data quality management should include explicit cultural bias checks in the creation and use of ontologies and knowledge graphs.

These challenges point to the necessity for a more inclusive, open, and mindful approach to the development and implementation of LOD systems, especially when working with cultural data. In conclusion, the panel underscored the importance of acknowledging and addressing cultural bias in Linked Open Data. By recognizing the inherent biases in existing systems and working towards more inclusive, transparent, and complex data representations, we can better serve the diverse needs of the Humanities and beyond. The proposed solutions emphasize the need for greater collaboration between computer scientists and humanities scholars, as well as a shift towards more inclusive practices in data collection, crowdsourcing, and tool development. Moving forward, it is crucial that the Semantic Web community continues to engage with these challenges, striving to create LOD systems that truly reflect the diverse range of perspectives and knowledge that constitute our shared cultural heritage.

Declaration on Generative AI

In the process of writing, we used generative AI technologies to improve the readability and language of the manuscript. The authors take full responsibility for the scientific content, insights, and conclusions presented in the work.

References

[1] O. Bruns, A. Poltronieri, L. Stork, T. Tietz (Eds.), Proceedings of the First International Workshop of Semantic Digital Humanities (SemDH 2024), co-located with ESWC 2024, volume 3724 of *CEUR Workshop Proceedings*, 2024. URL: https://ceur-ws.org/Vol-3724/.