

SSHOC-NL KG: Towards a Knowledge Graph for Social Sciences and Humanities

Andre Valdestilhas^{1,†}, Ronald Siebes^{1,†} and Jacco van Ossenbruggen^{1,†}

¹Vrije Universiteit Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands

Abstract

Social Sciences and (Digital) Humanities (SSH) increasingly benefit from the online availability of datasets, services, and tools provided by their peers and governmental organizations. However, the fragmented and heterogeneous nature of the metadata makes it difficult to assess the complexity of replicating results and the potential reuse of various parts (methodology, tools, and data) for follow-up research. For example, a significant effort is required to check the access requirements and quality of the data and to check if the format is compatible with tools that a researcher is familiar with. This paper outlines the vision, structure, and objectives of the Social Science and Humanities Open Cloud for the Netherlands (SSHOC-NL), focusing on its role in building and leveraging a national SSH knowledge graph. By developing a unified data environment, semantic integration tools, and interactive Data Stories, this SSHOC-NL effort transforms disparate data into interconnected scientific knowledge. This work aims to equip Dutch researchers with the tools and skills necessary to navigate complex societal challenges through a rich, semantically linked data landscape, representing a critical step towards realizing the promise of the European Open Science Cloud (EOSC) within the Dutch SSH domain.

Keywords

Social Sciences, Humanities, Knowledge Graph, SSHOC-NL

1. Introduction

The proliferation of digital data has catalyzed a profound transformation in the Social Sciences and Humanities (SSH). Researchers are no longer limited to traditional qualitative and quantitative methods but are increasingly leveraging large-scale, complex, and often sensitive datasets to investigate and historical phenomena with unprecedented depth and scale [1]. This data-driven paradigm, however, depends critically on the availability of robust, accessible, and well-managed research infrastructures capable of transforming raw data into structured, interconnected knowledge. The concept of a knowledge graph, which represents entities, their attributes, and their relationships in a structured format, including a scalable, effective, and user-friendly interface, offers a powerful approach to achieving this integration and enhancing scientific discovery.

In line with global movements towards open science, the FAIR guiding principles - which state that data should be findable, accessible, interoperable and reusable - have become the cornerstone of modern scientific data management [2]. Adherence and assessment [3] to these principles is essential for ensuring transparency, reproducibility, and the cumulative nature of scientific inquiry, all of which are prerequisites for building effective knowledge graphs. Furthermore, the European Open Science Cloud (EOSC) initiative seeks to federate existing research data infrastructures to provide a seamless environment for data storage, management, analysis, and reuse across disciplines and borders [4], with knowledge graphs emerging as a key enabler for semantic interoperability within such federated systems.

Despite these ambitions, the Dutch SSH data landscape is characterized by significant fragmentation and heterogeneity, which impedes the construction and utilization of a unified knowledge graph. To

ISWC 2025 Companion Volume, November 2–6, 2025, Nara, Japan

*Corresponding author.

[†]These authors contributed equally.

✉ a.valdestilhas@vu.nl (A. Valdestilhas); r.m.siebes@vu.nl (R. Siebes); jacco.van.ossenbruggen@vu.nl (J. v. Ossenbruggen)

ORCID 0000-0002-0079-2533 (A. Valdestilhas); 0000-0001-8772-7904 (R. Siebes); 0000-0002-7748-4715 (J. v. Ossenbruggen)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

address this, the Social Science and Humanities Open Cloud for the Netherlands (SSHOC-NL) was conceived.

This paper details the rationale, design, and demonstration of the SSHOC-NL Knowledge Graph. We show the potential of Linked Data schemas and vocabularies, semantically enriching and connecting research output, including papers, tools, and research data, integrated national infrastructure designed to facilitate the creation and exploitation of a comprehensive knowledge graph to empower the next generation of SSH research in the Netherlands.

2. The Knowledge Graph and Data Stories

The website kg.odissei.nl serves as a central hub for the SSHOC-NL knowledge graph. This platform is designed to house and present data stories, making complex social science data, publications, projects and tool more accessible, connected and understandable.

The kg.odissei.nl website currently functions as a container for **22 data stories**, which are built upon **23 datasets** and utilize **242 queries**. These numbers reflect the current scale of the knowledge graph, which aims to integrate disparate data sources to facilitate comprehensive research.

2.1. Schemas, vocabularies and graphs

To illustrate how we utilize schemas, vocabularies, and graphs to enhance interoperability, our Knowledge Graph (KG) integrates well-known and domain specific schemas and vocabularies, including the Simple Knowledge Organization System (SKOS)¹, the Bibliographic Ontology (BIBO)², and DCMI Metadata Terms³.

A practical example of a query - available online at <https://kg.odissei.nl/odissei/-/queries/A-toolsClariahOdissei/9> - demonstrates a real case that utilizes common vocabulary from Schema.org and DCMI Metadata Terms. This query retrieves information from two distinct graphs: the Clariah project⁴ and the ODISSEI code library⁵. All schemas, vocabularies, and graphs used in our KG are available online on our previously introduced website.

2.2. Data Stories

The data stories showcased on the platform demonstrate how various sources of information can be combined to answer complex research questions. Here are three examples:

- **The SSHOC-NL Knowledge Graph⁶**: This particular data story provides an in-depth overview and a clearer understanding of the type of information available within the graph. It highlights how queries can be constructed to combine otherwise isolated information sources, thereby revealing new insights. This story serves as an excellent introduction to the power and potential of the SSHOC-NL knowledge graph itself.
- **Exploring the Media Content Analysis Lab (MCAL) dataset⁷**: This data story focuses on the systematic analysis of large corpora of digital media content. This includes various forms of media from the Netherlands, such as traditional outlets (television, newspapers), new media (online news outlets), and social media platforms (Twitter, Instagram, Facebook). The MCAL dataset enables researchers to delve into the nuances of media representation and discourse.

¹<http://www.w3.org/2004/02/skos/core>

²<https://dcmi.github.io/bibo/>

³purl.org/dc/terms

⁴<https://tools.clariah.nl/data/>

⁵<https://w3id.org/odissei/ns/kg/graph/codelib/cbs>

⁶<https://kg.odissei.nl/odissei/-/stories/ODISSEI-Knowledge-Graph-the-story>

⁷<https://kg.odissei.nl/odissei/-/stories/mcal-basic-data-overview-story>

- **Tutorial SSHOC-NL⁸**: This data story aims to introduce knowledge graphs to members of the ODISSEI and SSHOC-NL projects. The goal is to improve understanding of the fundamental concepts, applications, and benefits of knowledge graphs in research. In this tutorial, we will explore how knowledge graphs can effectively organize and represent complex data relationships, allowing for more efficient data analysis and insights. Whether you are new to the concept or looking to deepen your understanding, this resource will serve as a valuable starting point in your journey with knowledge graphs.

The primary objective of the SSHOC-NL platform is to simplify the process of finding, accessing, reusing, and fully exploiting the potential of available information. By addressing the challenge of isolated data through the abstract concept of a knowledge graph, SSHOC-NL aims to foster more interconnected and impactful social science research.

2.3. The proof of concept

The SSHOC-NL Knowledge Graph (KG) is a prototype designed to explore the potential of knowledge graphs to establish future research infrastructures. It serves as a structured repository that connects various pieces of information, aiming to facilitate research and data sharing within the ODISSEI ecosystem.

The data integrated into the ODISSEI KG originates from diverse sources, as illustrated in fig. 1. This includes information on research projects conducted within the CBS⁹ remote access environment, research articles that have resulted from these projects, and research software developed for the LISS¹⁰ and CBS¹¹ projects. Furthermore, all data sets accessible through the Dataverse API¹² are incorporated into the knowledge graph, ensuring a comprehensive collection of relevant research data and the main concepts illustrated in fig. 2, also available online¹³.

The organization of the ODISSEI KG is guided by several key design principles, including the reuse of existing linked data standards and persistent identifiers. For instance, SKOS is utilized for vocabularies and Dublin Core for publication metadata, while persistent identifiers such as DOIs for papers and datasets, ORCIDs for authors, and RORs for institutes are employed. Data are processed and organized through an ETL (extract, transform, and load) pipeline, which uses specific scripts to ingest and transform raw data into a linked data format, and this pipeline is managed via a GitLab CI/CD system¹⁴.

More information about the SSHOC-NL Knowledge Graph, such as documentation, queries, experiments reproduction, and data stories, is available at <https://kg.odissei.nl/>.

3. Discussion

The development of SSHOC-NL, with its central focus on a national knowledge graph, represents a transformative step for the Dutch research landscape. It is designed from the ground up to be a key component of the European Open Science Cloud. By adopting international semantic standards and fostering interoperability, SSHOC-NL aims to ensure that Dutch researchers can contribute to and benefit from the broader European knowledge ecosystem. However, the success of the project will depend not only on the technology but also on the active participation of the research community. The focus on training in knowledge graph technologies and community building is therefore critical for embedding semantic web principles and Open Science practices into the culture of SSH research.

⁸<https://kg.odissei.nl/odissei/-/stories/Tutorial-SSHOC-NL>

⁹<https://www.cbs.nl/>

¹⁰<https://portal.odissei.nl/dataverse/liss>

¹¹<https://portal.odissei.nl/dataverse/cbs>

¹²<https://dataverse.nl/>

¹³<https://github.com/odissei-data/odissei-kg/blob/acceptance/static/mainConcepts.ttl>

¹⁴publicly available in GitHub as well at <https://github.com/odissei-data/odissei-kg>

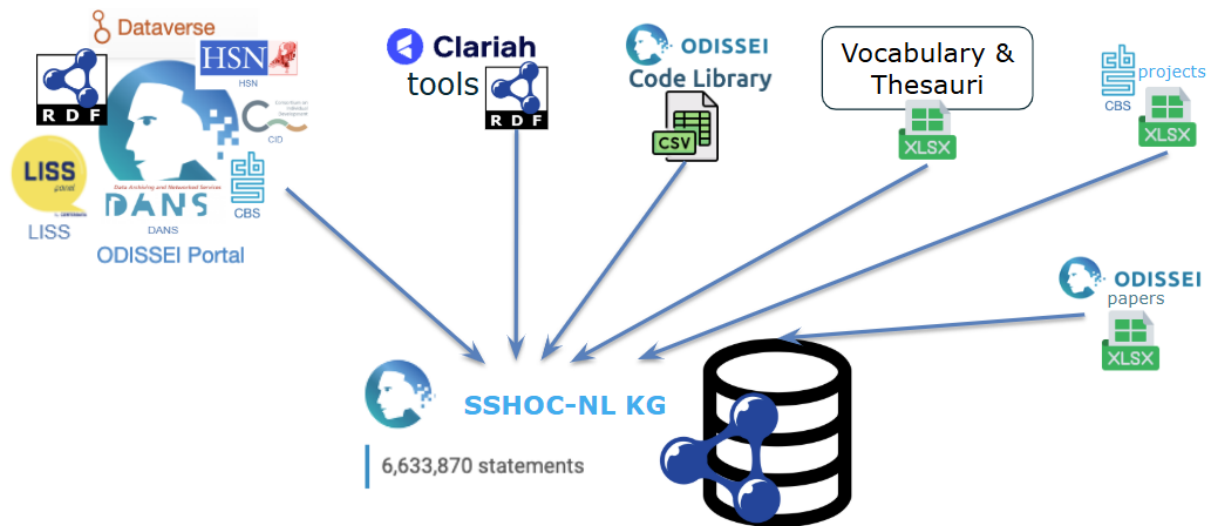


Figure 1: Metadata sources included in the Knowledge Graph.

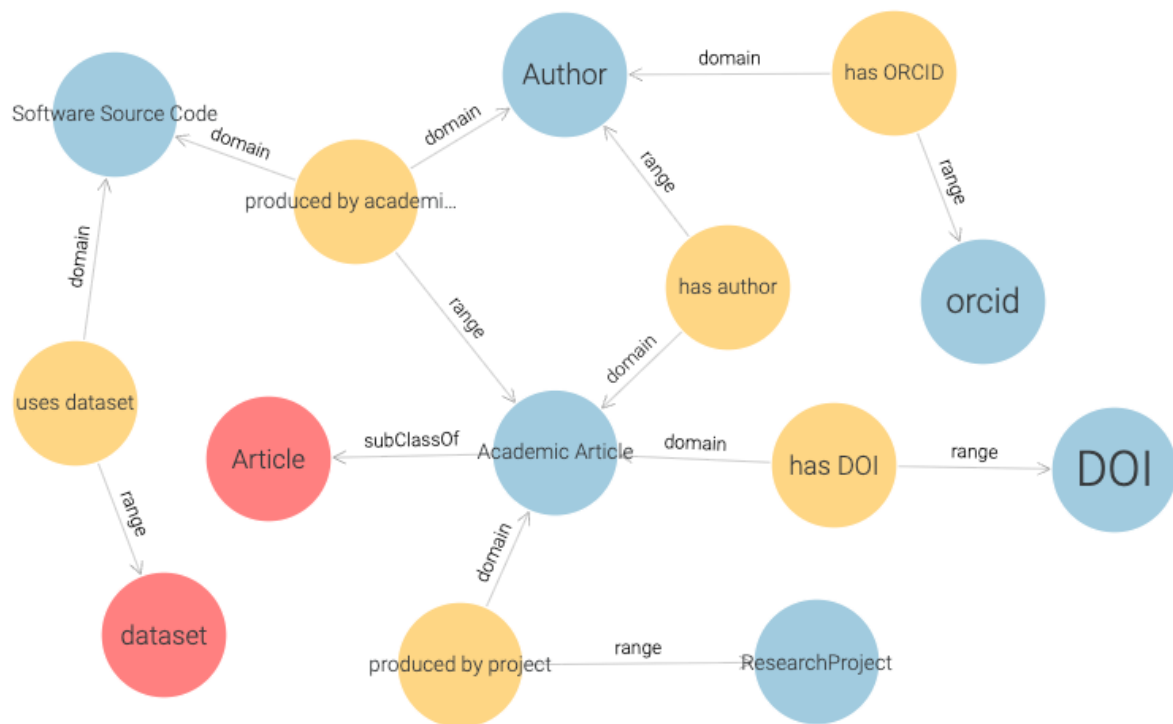


Figure 2: A simplified version of the main concepts present in one of our DataStories available at <https://kg.odissei.nl/odissei/-/stories/ODISSEI-Knowledge-Graph-the-story>

4. Conclusion

SSHOC-NL is a timely and essential investment in the future of the Social Sciences and Humanities in the Netherlands, specifically through its commitment to building a national knowledge graph. It directly addresses the critical bottleneck of data fragmentation and semantic disconnect by delivering a unified, secure, and user-centric cloud infrastructure centered around interconnected knowledge. Through its integrated portal, semantic integration efforts, and demonstrations, we hope that the SSHOC-NL KG will provide researchers with unprecedented access to the nation's rich SSH data resources, tools, and

research output. Our future plans include the use of the Croissant metadata format¹⁵, integration with ORKG¹⁶, OpenCitations¹⁷, and the SODA code¹⁸. Additional information is available on the project website <https://sshoc.nl/>.

Declaration on Generative AI

During the preparation of this work, the authors utilized Grammarly to correct and spell-check, as well as to improve the text's grammatical readability. After using the tool, the authors reviewed and edited the content as needed to take full responsibility for the publication's content.

References

- [1] C. L. Borgman, Big data, little data, no data: Scholarship in the networked world, MIT press, 2015.
- [2] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, et al., The fair guiding principles for scientific data management and and stewardship, *Scientific Data* 3 (2016) 1–9.
- [3] A. Valdestilhas, M. Windhouwer, R. Siebes, S. Wang, Comparing fair assessment tools and their alignment with fair implementation profiles using digital humanities datasets, in: *International Workshop of Semantic Digital Humanities*, 2025.
- [4] European Commission, Turning FAIR into reality: Final report and action plan from the European Commission expert group on FAIR data, Technical Report, Publications Office of the European Union, 2018.

¹⁵<https://docs.mlcommons.org/croissant/docs/croissant-spec.html>

¹⁶<https://orkg.org/>

¹⁷<https://opencitations.net/>

¹⁸<https://fair.odissei-soda.nl/>