DCPAPERS

DCMI 2026

Proceedings of the International Conference on Dublin Core and Metadata Applications 2026

ISSN: 1939-1366

Metadata-Driven Semantic Interoperability: The HerStory-NeSyAI project for Trustworthy Neurosymbolic AI in Digital Humanities

Miquel Centelles[1], **Matheus Jenevain**[1], **Elena Gómez**[1], and **Núria Ferran-Ferrer**[1]

[1] Universitat de Barcelona

Abstract

This short paper presents *HerStory-NeSyAI*, a work-in-progress Digital Humanities project that uses metadata-driven semantic interoperability to connect fragmented historical datasets. The project addresses fragmented historical datasets through metadata-driven semantic interoperability, combining a knowledge graph, ontology layer, and retrieval-augmented generation (RAG). It treats interoperability as a condition for epistemic justice and explores how graph-grounded metadata can improve transparency, traceability, and data integrity while mitigating bias, hallucinations, and poisoning risks. In operational terms, these goals are implemented through the project's prototype infrastructure, which supports source-grounded querying and a three-path strategy for connecting heterogeneous datasets through semantic integration, semantic linking, or MCP/API-mediated access.

1. Introduction

Digital Humanities datasets are often abundant yet structurally fragile. Many are created in short-term projects, remain dependent on unstable portals, and are described with uneven metadata, making them difficult to find, verify, connect, and reuse over time. This creates silos: datasets are not necessarily lost, but they become inaccessible or functionally disconnected from wider scholarly and public knowledge ecosystems. In sensitive historical domains, such fragmentation also has representational consequences, since already marginalized people and groups are more likely to remain isolated across collections and therefore less visible in Artificial Intelligence (AI)-assisted access environments [1], [2].

HerStory-NeSyAI[1] project addresses this problem in the context of datasets on Francoist repression and censorship in Spain, with a gender and intersectional perspective and with explicit attention to interoperability with Wikidata and related open knowledge ecosystems [3]. Rather than treating metadata as a descriptive afterthought, the project treats metadata, ontology, provenance, and validation as operational infrastructure for explanation, reuse, and governance.

A clear distinction is needed between *HerStory-NeSyAI* and the project's prototype infrastructure. *HerStory-NeSyAI* is the research project: it defines the problem space, the epistemic-justice rationale, and the methodological orientation. The project's prototype infrastructure is the operational environment through which heterogeneous historical datasets are semantically processed, connected to a governed knowledge

[1]

Short Paper

Available Aug 01, 2026

Correspondence to
Miquel Centelles
miquel.centelles@ub.edu

graph, and queried through source-grounded conversational access. In other words, *HerStory-NeSyAI* provides the conceptual and research framework, while the prototype infrastructure provides the environment in which that framework is implemented and tested.

The contribution of this paper is therefore not a generic 'Knowledge Graph + Retrieval Augmented Generation' proposal. Its novelty lies in combining a values-aligned framing of metadata and epistemic justice with a practical interoperability rule for fragmented historical datasets. The paper makes four contributions. First, it reports the current state of *HerStory-NeSyAI* as a work-in-progress project. Second, it argues that epistemic justice is a design principle for metadata and interface decisions, not a later ethical add-on. Third, it shows why graph-grounded retrieval is a plausible route for improving transparency, traceability, and data integrity in AI-assisted historical research. Fourth, it proposes a three-path strategy, implemented through the prototype infrastructure, for deciding how heterogeneous datasets should be connected according to their condition and data model.

2. Epistemic justice as conceptual foundation

In *HerStory-NeSyAI*, epistemic justice is not a final evaluative layer but the conceptual basis of the infrastructure itself [4], [5]. Semantic and neurosymbolic systems for cultural heritage should therefore not be assessed only in terms of usability or retrieval performance. They should also be evaluated in terms of trust, accountability, participation, representational fairness, provenance, and governance. When these safeguards are absent, cultural heritage systems risk reproducing testimonial and hermeneutical injustices by including marginalized communities as data sources while excluding them from interpretive control [4].

This perspective is especially relevant where women and other minority gender identities have already been marginalized in historical sources and are more vulnerable to invisibility in databases and AI-mediated interfaces. *HerStory-NeSyAI*, therefore treats semantic linkage, metadata governance, and traceable representation as prerequisites for fairer access [5]. In this framing, interoperability is not only a technical convenience. It is also a way of reducing epistemic asymmetries by making scattered entities, events, and documents discoverable across collections, and by grounding AI outputs in explicit evidence structures rather than opaque statistical generation alone.

The project also assumes that trust in semantic and neurosymbolic systems is a socio-technical outcome grounded in meaningful participation and accountability [6]. This is why *HerStory-NeSyAI* links technical architecture to co-creation, provenance, and governance from the outset [7]. More broadly, this stance is compatible with community-centered stewardship approaches such as the CARE Principles, which emphasize collective benefit, authority to control, responsibility, and ethics when working with marginalized or sensitive data [8].

3. The *HerStory-NeSyAI* project approach and the role of the prototype infrastructure

HerStory-NeSyAI adopts a hybrid approach that couples a large language model with a knowledge graph and ontology layer used as the knowledge base for retrieval-augmented generation (RAG) [9]. The current design combines an ontology layer, an entity instance layer, semantic reconciliation with Wikidata-oriented identifiers, and graph-based retrieval intended to support grounded and explainable outputs [3]. The metadata layer is currently Wikidata-oriented, but the project also treats crosswalks to DCMI Terms, DCAT, and schema.org as a practical requirement for moving historical records across repositories, interfaces, and reuse contexts. In this sense, it follows a neurosymbolic model in which retrieval, inference, generation, and governed updates remain tightly connected [9].

The prototype infrastructure is the operational semantic and conversational environment through which this approach is instantiated. It supports four concrete functions already central to the project: (1) ingestion and semantic mapping of heterogeneous historical datasets, (2) entity reconciliation and alignment with governed identifiers, (3) graph-based querying and coverage audits, and (4) conversational access grounded in retrieved graph evidence rather than unstructured text alone [10]. A historian or curator should be able to query the graph in natural language, inspect the linked entities and their provenance, and move from answers to supporting records. This means that the prototype is not only a retrieval layer but also an interface for semantic inspection, coverage checking, and accountable reuse. Operationally, the prototype infrastructure links three interdependent layers. First, heterogeneous source datasets are ingested, normalized, and semantically aligned through mapping and reconciliation workflows. Second, the resulting entities, relations, and metadata are organized in an ontology-driven knowledge graph layer that supports linked evidence and reusable semantic structures. Third, retrieval components use this graph as the evidentiary basis for source-grounded querying, so that conversational outputs can be traced back to linked entities, provenance statements, and documented connection choices. The distinction is therefore clear: *HerStory-NeSyAI* defines the research problem, conceptual framework, and methodological orientation, whereas the prototype infrastructure provides the environment through which those principles are implemented and tested.

The broader institutional framework is the University of Barcelona's effort to move from isolated datasets toward a connected Digital Humanities infrastructure built on three pillars: standardization and FAIR/Linked Open Data compliance, knowledge-graph integration supported by neurosymbolic AI, and participatory human-centered design. At this stage, both the project and the infrastructure should be understood as early methodological work rather than finalized systems. Their main value lies in showing how metadata-rich, ontology-guided, and provenance-aware infrastructures can shape more reliable AI-assisted access to humanities collections.

A short clarification on implementation status is useful. Built today are semantic ingestion and mapping workflows, entity reconciliation procedures, graph-based retrieval, and source-grounded querying over selected datasets. Planned next are broader MCP-based connectors, more systematic white-box logging of agent actions, and structured evaluation of how far the prototype mitigates hallucination, bias, and poisoning risks in practice.

4. Graph-grounded metadata against bias, hallucinations, and poisoning

A central working hypothesis of the project is that graph-grounded metadata can reduce important generative-AI risks better than text-only pipelines [11]. The approach is designed so that the model does not rely on diffuse internal memory alone, but retrieves evidence from a knowledge graph where nodes, relations, and sources can be identified and inspected. This makes it possible to connect outputs to explicit semantic structures and traceable sources, which improves explainability and transferability.

This matters first for **bias**. *HerStory-NeSyAI* is explicitly framed around the need to address the invisibility of women and other marginalized groups in historical databases and AI systems [5], [12]. By aligning heterogeneous sources, reconciling entities, and using ontology-guided structures, the system seeks to reduce some of the distortions produced when isolated or weakly contextualized records are consumed by generic AI models [13].

It also matters for **hallucination**. Earlier project materials identify hallucination as a major limitation of current Large Language Model (LLM)-based interaction and present hybrid AI as a more explainable alternative because responses can be tied to graph evidence rather than generated without accountable grounding. Recent work on knowledge-graph-supported mitigation strengthens this rationale [14].

Finally, it matters for **poisoning and integrity risks**. The project's internal reflections on poisoning emphasize that open and weakly traceable AI pipelines remain vulnerable to manipulation, whereas curated symbolic structures create a more governable evidentiary substrate. The claim here is not that a knowledge graph eliminates hallucination, but that it creates a more inspectable evidentiary basis. This logic is consistent with the AI Act's emphasis on data governance, technical documentation, transparency, record-keeping, human oversight, and traceability in higher-risk AI contexts [15], [16].

5. From data silos to interoperable humanities infrastructures

A distinctive contribution of the project is its explicit methodology for deciding how a dataset should connect to ***the prototype***. The decision is based on two questions: **what is the dataset's health status** and **what is its data model**. This classification determines one of three connection types.

5.1. Type 1: Semantic integration into the knowledge graph

This route is used for fragile or at-risk datasets, especially those dependent on unstable portals or tending toward fossilization. It is the rescue path. A concrete example is the *Cultura y censura* database. Because it is historically valuable, structurally heterogeneous, and central to the Francoist censorship case, it is best treated as a Type 1 dataset: the priority is semantic rescue and integration. In practice, this means mapping its records into a governed graph, reconciling persons and works with Wikidata-oriented identifiers, and making them queryable together with provenance and contextual metadata [3]. The point of the example is that the three-path model is not determined by topic alone, but by the combination of preservation conditions and data model. By contrast, datasets that are already stably preserved and reusable, but remain non-RDF, are better handled through the Type 3 route rather than through full semantic integration.

5.2. Type 2: Semantic linking with the knowledge graph

This route is used for robust datasets that are already expressed in RDF. In these cases, the dataset does not need to be absorbed into the central graph. Instead, it is semantically linked to the graph and queried in a federated way, typically through SPARQL endpoints. This avoids unnecessary duplication and preserves local stewardship while enabling cross-dataset retrieval and explanation [3].

5.3. Type 3: Model Context Protocol (MCP) or Application Programming Interface (API) connection

This route is used for **robust datasets that are not in RDF**, such as relational or document-oriented systems. Here the proposal is to develop an MCP layer connected to a federated-access API so that multiple heterogeneous sources can be queried in a uniform and reusable way without centralizing the data. In practical terms, MCP is not the same thing as the federated-access middleware it wraps. Rather, it is the protocol-facing layer through which the model requests an operation, while the underlying infrastructure may rely on tools such as Trino for structured databases and SPARQL for knowledge graphs. In this design, MCP does not replace semantic modeling; it extends the infrastructure to robust non-RDF datasets that still require controlled and auditable AI access.

This three-path strategy matters because it turns interoperability into a governed methodological choice rather than a one-size-fits-all technical solution [17]. It allows preservation, linkage, and AI-assisted access to be adapted to the actual condition of the dataset, while keeping the process explainable and auditable [18].

6. Trustworthy AI, transparency, and public value

The project's broader claim is that trustworthy AI in Digital Humanities depends on metadata, provenance, and governance as much as on model capability [19]. The value of the prototype infrastructure is therefore not simply that it can answer questions in natural language, but that it can do so through evidence-bearing structures,

documented connection choices, and reusable semantic artifacts. This also helps frame humanities datasets as a public good: preserving, connecting, and explaining them increases their scholarly, cultural, and civic value beyond the lifespan of any individual project portal.

In this respect, the MCP proposal is also relevant to transparency. A white-box architecture can record the trace of the agent's reasoning and execution process, including actions taken, parameters used, and generated justifications, so that the operation of the system remains inspectable rather than opaque [20]. Where sensitive data are involved, irreversible anonymization can be applied before data leave their original source, allowing federated access while reducing re-identification risks. These measures are coherent with current orientations toward transparency, logging, documentation, and human oversight [16].

7. Conclusion

HerStory-NeSyAI is currently best understood as a **work-in-progress project and methodological proposal** for connecting metadata, knowledge graphs, and AI in a way that is technically robust and normatively grounded. Its main contribution is to show that semantic interoperability can be treated as a form of epistemic justice: a means of making fragile and disconnected humanities data more discoverable, more linkable, more reusable, and more accountable in AI-assisted access environments.

Within this framework, the prototype infrastructure provides the operational environment through which these principles are operationalized and tested. Their combined contribution is therefore twofold: a conceptual and methodological orientation grounded in epistemic justice, and an infrastructural strategy for implementing it through governed semantic interoperability. Together, they point toward a more explainable, auditable, and socially responsible infrastructure for Digital Humanities.

Acknowledgements

This research was funded by the Spanish Ministry of Science, Innovation and Universities under the project *Connecting women's history to Neuro-Symbolic AI: An Information Architecture approach to Francoist repression case study (HerStory&NeSyAI)*, grant **PID2023-147673OB-I00**.

References

[1] T. Gebru, "Race and gender," in *The Oxford Handbook of Ethics of AI*, 1st ed., M. D. Dubber, F. Pasquale, and S. Das, Eds., Oxford University Press, 2020, pp. 252–269. doi: 10.1093/oxfordhb/9780190067397.013.16.

[2] N. Ferran-Ferrer, J.-J. Boté-Vericad, and J. Minguillón, "Wikipedia gender gap: A scoping review," *El Profesional de la Información*, vol. 32, no. 5, p. e320617, 2023, doi: 10.3145/epi.2023.nov.17.

[3] M. Centelles Velilla and N. Ferran-Ferrer, "Using Wikidata's ontology in practice: A neuro-symbolic, community-centred workflow for integrating and reusing humanities datasets," *Journal of Open Humanities Data*, vol. 11, p. 84, 2025, doi: 10.5334/johd.439.

[4] M. Fricker, *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford University Press, 2007.

[5] N. Ferran-Ferrer and M. Centelles Velilla, "From silence to spotlight: Towards epistemic justice in women's digital representation in the age of Digital Humanities and Artificial Intelligence," in *Why EDI Matters: Equ(al)ity, Diversity and Inclusion in European Universities*, A. Oboe, V. Sokolová, and H. W. Henriksson, Eds., Padua University Press, 2025, pp. 29–46. doi: 10.25430/pupb-2025-9788869384820.

[6] F. Delgado, S. Yang, M. Madaio, and Q. Yang, *The participatory turn in AI design: Theoretical foundations and the current state of practice*. in Equity and Access in Algorithms, Mechanisms, and Optimization. ACM, 2023, pp. 1–23. doi: 10.1145/3617694.3623261.

[7] S. Jasanoff, "The idiom of co-production," in *States of Knowledge: The Co-Production of Science and the Social Order*, Routledge, 2004, pp. 1–12. doi: 10.4324/9780203413845.

[8] S. R. Carroll *et al.*, "The CARE principles for Indigenous data governance," *Data Science Journal*, vol. 19, no. 1, p. 43, 2020, doi: 10.5334/dsj-2020-043.

[9] P. Hitzler, F. Bianchi, M. Ebrahimi, and M. K. Sarker, "Neural-symbolic integration and the Semantic Web," *Semantic Web*, vol. 11, no. 1, pp. 3–11, 2020, doi: 10.3233/SW-190368.

[10] V. de Boer, *Knowledge graphs for cultural heritage and Digital Humanities*. in SUMAC@ACM Multimedia. 2023, p. 3. [Online]. Available: https://www.victordeboer.com/wp-content/uploads/2023/11/SUMAC_keynote1_compressed-1.pdf

[11] G. Agrawal, T. Kumarage, Z. Alghamdi, and H. Liu, *Can knowledge graphs reduce hallucinations in LLMs? A survey*. in Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers). Association for Computational Linguistics, 2024, pp. 3947–3960. doi: 10.18653/v1/2024.naacl-long.219.

[12] A. Fan and C. Gardent, *Generating biographies on Wikipedia: The impact of gender bias on the retrieval-based generation of women biographies*. in Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Association for Computational Linguistics, 2022, pp. 8561–8576. doi: 10.18653/v1/2022.acl-long.586.

[13] C. C. Zhang and L. Terveen, *Quantifying the gap: A case study of Wikidata gender disparities*. in 17th International Symposium on Open Collaboration. ACM, 2021, pp. 1–12. doi: 10.1145/3479986.3479992.

[14] X. Guan and et al., "Mitigating large language model hallucinations via autonomous knowledge graph-based retrofitting," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, no. 16, pp. 18126–18134, 2024, doi: 10.1609/aaai.v38i16.29770.

[15] European Union, "Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) (Text with EEA relevance)," 2024. [Online]. Available: https://data.europa.eu/eli/reg/2024/1689/oj

[16] Future of Life Institute, "High-level summary of the AI Act," EU Artificial Intelligence Act. [Online]. Available: https://artificialintelligenceact.eu/high-level-summary/

[17] D. Abián, A. Meroño-Peñuela, and E. Simperl, *An analysis of content gaps versus user needs in the Wikidata knowledge graph*. in The Semantic Web – ISWC 2022, no. 13489. Springer International Publishing, 2022, pp. 354–374. doi: 10.1007/978-3-031-19433-7_21.

[18] M. Miquel-Ribé and D. Laniado, "The Wikipedia Diversity Observatory: Helping communities to bridge content gaps through interactive interfaces," *Journal of Internet Services and Applications*, vol. 12, no. 1, p. 25, 2021, doi: 10.1186/s13174-021-00141-y.

[19] B. Shneiderman, "Bridging the gap between ethics and practice: Guidelines for reliable, safe, and trustworthy human-centered AI systems," *ACM Transactions on Interactive Intelligent Systems*, vol. 10, no. 4, pp. 1–31, 2020, doi: 10.1145/3419764.

[20] A. Mileo, "Towards a neuro-symbolic cycle for human-centered explainability," *Neurosymbolic Artificial Intelligence*, vol. 1, pp. 1–12, 2025, doi: 10.3233/nai-240740.