Authoritative Practices and Collective Validation: Wikidata within the Collaborative Digital Edition of the Greek Anthology

Mathilde Verstraete Maxime Guénette Marcello Vitali-Rosati

Introduction

The management and preservation of research data in the Humanities increasingly raises questions about its sustainability, sharing, and validation. In this context, Wikidata constitutes a powerful and collaborative tool. By challenging traditional models where researchers act as both producers and gatekeepers of authority, Wikidata redefines these issues and fosters new paradigms of collaboration.

> ❗ To add
>
> Données issues de la recherche comme golden standard -> problématique

Context

Since 2014, Marcello Vitali-Rosati and his team have been developing the project for a (digital and collaborative) edition of the Greek Anthology (Mellet, 2020; Vitali-Rosati et al., 2021; Mellet and Vitali-Rosati, 2021). The project arose from a need to index and make accessible this vast corpus, a monument of Greek literature. Thanks to the work of numerous contributors with varied profiles, the project's publication platform presents - for each epigram - its location in the Palatine manuscript (the Codex palatinus gr. 23, the main manuscript providing us with the Palatine Anthology), which we have retrieved from the API of the Heidelberg Library's manuscript annotation tool using the IIIF protocol, several translations into different languages, information about the author of the epigram and keywords for the themes that make it up, comments, internal and external references, and alignments between the different translations. A REST API has been set up to query our data.

In this paper, we focus on keywords, which include authors, cities, and several other categories such as collections, divinities, epithets, and epiclesis (a full list is available at

https://anthologiagraeca.org/keywords/). In the course of the project, we implemented a rule requiring all keywords to be linked to Wikidata, meaning that any new keyword added must include a corresponding Wikidata identifier. This decision led to a systematic effort to reconcile our existing data with Wikidata.

Wikidata thus plays a central role in the AG project. All keywords used to annotate the platform either have a Wikidata identifier or are created accordingly – note that some old keywords still have not been linked to a wikidata id yet, we will talk about that *infra*. This integration proved particularly useful in addressing inconsistencies within our list of authors. Since both Wikidata and our data model are multilingual, discrepancies such as missing authors, duplicate entries, and inconsistent information across languages were directly reflected on our platform (https://anthologiagraeca.org/authors/). To resolve these issues, we first ensured that all authors on the platform had a Wikidata URN. We then expanded our contribution by searching for author names in multiple languages (French, English, Italian, Ancient Greek, and Latin) and adding them to Wikidata. Once this information was uploaded, the Wikidata community quickly reviewed and refined our data to align it with their standards. This process not only enhanced our own dataset but also strengthened Wikidata's overall accuracy and consistency.

! To add

- AG
- Chantier Auteurs
- Interactions avec la communauté

Quelques problématiques

On co-existing informations and visions du monde

! To add

- Coexistence des informations ?
- Diversité des visions du monde ?

  peut-être exemplifier avec le label officiel pour les alternatives label ? (genre, "je veux que ma data soit le vrai label officiel et pas relégué à un label alternatif")

Wikidata as a multilingual authority

On delegating the curation of academic data

As this example shows, because we have chosen to involve a wider community as an authority figure, the AG project is no longer the sole custodian of its data. By delegating part of the curatorial process to Wikidata and its contributors, we embrace a model of distributed authority that challenges traditional academic hierarchies. This delegation implies a shift not only in who validates the data, but also in how validation itself is understood — no longer as a top-down, expert-driven process, but as a collective negotiation of meaning, relevance, and accuracy.

This shift towards distributed authority has many implications.

What are the implications of this shift toward distributed authority. How can that shift in authority benefit academic research projects? Is Wikidata's epistemological paradigm coherent with ours? Can we think of a generic epistemological framework to be effectively applied to specific academic endeavors?

Wikidata and Digital Humanities

Since its creation in 2012, Wikidata have progressively become one of the most important knowledge graph on the web. As structured data are increasingly central to the organization, retrieval, and circulation of information online, Wikidata plays a pivotal role in shaping how knowledge is represented, accessed, and reused.

For this reason, Wikidata is growing in popularity in Digital Humanities (DH). Although there was — and to some extent still is — skepticism regarding its quality and potential as a research tool or platform, recent studies have shown that Wikidata is now widely used across DH projects.[1] It is primarily seen in this discipline as a content provider to access, publish or disseminate Linked Open Data (LOD) without the technical and financial barriers of the Semantic Web.

For others, such as GLAM (Galleries, Libraries, Archives and Museums) institutions, Wikidata is rather used as a publishing platform and metadata curation. As digitization and digital strategies have usually been independent from one institution to another, there is now different levels of accessibility and discoverability in the GLAM sector.[2] Thus, Wikidata is

[1] Stacey Cook, "The Uses of Wikidata for Galleries, Libraries, Archives and Museums and Its Place in the Digital Humanities," *Comma* 2017, no. 2 (2019): 117–24, https://doi.org/10.3828/comma.2017.2.12; Fudie Zhao, "A Systematic Review of Wikidata in Digital Humanities Projects," *Digital Scholarship in the Humanities* 38, no. 2 (2023): 852–74, https://doi.org/10.1093/llc/fqac083.

[2] Alicia Fagerving, "Wikidata for Authority Control: Sharing Museum Knowledge with the World," *Digital Humanities in the Nordic and Baltic Countries Publications* 5, no. 1 (2023): 222–39, https://doi.org/10.5617/dhnbpub.10665.

used either as a publication platform to create for the first time a digital identifier for cultural heritage enriched with LOD or as a platform to better disseminate cultural collection, link their institutional platform to Wikidata and enrich their metadata.[3]

Beyond its function as a repository or a platform for metadata curation, Wikidata also serves a more connective role. Its structure as a knowledge graph and its integration with Linked Open Data principles enable it to act as a bridge between otherwise siloed datasets. By assigning persistent identifiers and encouraging alignment across vocabularies and standards, Wikidata facilitates interoperability between diverse projects and institutions.

Wikidata as a linking hub

Thus, according to DH projects and GLAM institutions, one of the main advantages of Wikidata is being a hub to link different external resources and datasets.[4] This can be done be creating external identifiers properties that will link entities of diverses databases to their correspondent Wikidata item.

Lets take for example Megara, an ancient Greek *polis* mentioned three times in the Greek Anthology (AG 7.124; 7.337; 12.129). Its Wikidata item (Q42307600) is linked to dictionaries, libraries catalogues and most importantly many databases specialized on Greco-Roman Antiquity such as Pleiades, ToposText or MANTO. By doing so, Wikidata is playing a pivotal role for easy cross-referencing informations and while becoming an authority control mechanism.[5] As Linked Open Data are meant to add context to data, it becomes much more efficient to enrich a dataset directly from Wikidata or by cross-querying different databases that are all linked to a single Q item.

Some even goes further and suggests that Wikidata should not be used as a centralized platform to link external resource identifiers, but should rather be used as "the" identifier.[6] Having too many identifiers for the same resource (like our example Megara) can be confusing et results in bad reconciliation over time. Using Wikidata as the universal identifier provider would then have many advantages such as having a single description model, offering single SPARQL endpoint and API for accessing and processing data or providing a sustainable option for data storing.

[3] Gustavo Candela et al., "A Systematic Review of Wikidata in GLAM Institutions: A Labs Approach," in *Linking Theory and Practice of Digital Libraries*, ed. Apostolos Antonacopoulos et al. (Cham: Springer Nature Switzerland, 2024), 34–50, https://doi.org/10.1007/978-3-031-72440-4_4.

[4] Joachim Neubert, "Wikidata as a Linking Hub for Knowledge Organization Systems? Integrating an Authority Mapping into Wikidata and Learning Lessons for KOS Mappings," in *Proceedings of the 17th European Networked Knowledge Organization Systems Workshop Co-Located with the 21st International Conference on Theory and Practice of Digital Libraries 2017 (TPDL 2017), Thessaloniki, Greece, September 21st, 2017*, ed. Philipp Mayr et al., vol. 1937, CEUR Workshop Proceedings (CEUR-WS.org, 2017), 14–25.

[5] Fagerving, "Wikidata for Authority Control."

[6] Theo Van Veen, "Wikidata: From 'an' Identifier to 'the' Identifier," *Information Technology and Libraries* 38, no. 2 (2019): 72–81, https://doi.org/10.6017/ital.v38i2.10886.

Although this position seems more radical, it does point to one other critical aspect of the relationship between Wikidata and external resources: reciprocal contributing. As demonstrated by a paper on the reciprocal relationship by Wikidata and VIAF, a international platform that combines multiples institutional name authority files, reciprocal comparison between external resources and Wikidata can improved data quality for each platform.[7] Contributions by GLAM institutions such as libraries could then be transferred to Wikidata and vice-versa.

However, there is an obvious difference in authority figure between a contribution from an institution and one from a Wikidata editor. In academia, authority is frequently synonymous with expertise and influence, recognized and exerted within specific discipline. Outside academia, governments or GLAM institutions are also seen as a source of authority. Within any given area of scholarship, there are individuals who have published influential works and who are regarded as authorities on the subjects they research and write about. This interpretation of the authoritative figure typically establishes a top-down hierarchy, positioning traditional producers of knowledge as its sole custodians.

We argue that this top-down hierarchy is contextual and is not always what is best for knowledge dissemination, especially fort digital platforms. As collaborative, crowdsourced, and peer-contributed platforms such as Wikidata increasingly establish themselves as new forms of authority both within and beyond academia, they raise important questions about the seal of authority and the reliability of the data they provide. This is primarily due to the fact that, in most cases, anyone can contribute to these platforms, which leads some to question the reliability and quality of the data.

Thus, the status of Wikidata's editors as an authority figure is questioned because they are mostly seen by scholars and professionals as "amateurs".[8] Despite the increasing involvement of non-professionals in cultural initiatives, particularly on digital platforms, their contributions are still often perceived as less authoritative and must be validated by professionals to be considered reliable. This dynamic creates tensions between new actors of knowledge creation and prior groups that still are ambiguous about how to integrate them.

In the case of Wikidata, this manifests in conflicting perceptions of authority: while the platform encourages collaborative knowledge production and open participation, its data is often deemed trustworthy only when curated or approved by recognized experts. Studies have shown that, while there is still room for improvement in Wikidata's data, the platform

[7] Carlo Bianchini, Stefano Bargioni, and Camillo Carlo Pellizzari Di San Girolamo, "Beyond VIAF: Wikidata as a Complementary Tool for Authority Control in Libraries," *Information Technology and Libraries* 40, no. 2 (2021), https://doi.org/10.6017/ital.v40i2.12959.

[8] We define "amateur" as a socially engaged non-professional actor who, outside formal institutional frameworks, contributes to cultural, artistic, or documentary activities—often via participatory digital platforms. This figure is usually defined by a high level of self-taught expertise, an individual pursuit of excellence, a paradoxical sociability built on both emancipation and community, and a privileged relationship with digital platforms as new spaces for recognition and knowledge production. See Marta Severo, *L'impératif participatif: Institutions culturelles, amateurs et plateformes* (Bry-sur-Marne: Institut National de l'Audiovisuel (INA), 2021), https://doi.org/10.3917/ina.sever.2021.01.

is increasingly recognized as a high-quality knowledge graph—one whose quality is context-dependent and must be assessed on a case-by-case basis[9] Its growing community actively contributes to the growing improvement of data quality, while control mechanisms such as ShEx Schemas are gradually being implemented to ensure that Wikidata items conform to data models collaboratively defined by the community.[10] These community-driven standards not only support data consistency but also reflect the platform's commitment to a decentralized yet structured form of knowledge governance.

To this end, one of the best ways to ensure the quality and relevance of Wikidata is not merely to assess it from the outside, but rather to contribute directly to it. Through engagement with the platform—making statements, creating data models, correcting mistakes, and arguing over ontologies—researchers, cultural organizations, and "amateurs" can directly shape the knowledge graph. Rather than viewing Wikidata as a finished product to be critiqued, it becomes an open and participatory space where quality is a byproduct of collective interaction.

This approach echoes with broader theories of collective intelligence. Collective intelligence refers to the shared or group intelligent derived from collaboration, collective work, and competition among thousands of actors working together within a common setting. Within collective intelligence, the wisdom of the crowd is considered superior to individual or isolated intelligences.[11] Platforms such as Wikidata embody the best case of how community decentralized work may lead towards generating structured quality knowledge—provided that the contributions are supported by transparent rules, mechanisms of coordination, and mutual trust among contributors.[12]

As Pierre Lévy argues, collective intelligence is not about homogenizing knowledge or erasing differences in expertise, but about creating a space where knowledge can circulate, evolve, and

[9] Alessandro Piscopo and Elena Simperl, "What We Talk about When We Talk about Wikidata Quality: A Literature Survey," in *Proceedings of the 15th International Symposium on Open Collaboration* (Skövde Sweden: ACM, 2019), 1–11, https://doi.org/10.1145/3306446.3340822; Kartik Shenoy et al., "A Study of the Quality of Wikidata," *Journal of Web Semantics* 72 (2022): 100679, https://doi.org/10.1016/j.websem.2021.100679; Zhao, "A Systematic Review of Wikidata in Digital Humanities Projects."

[10] On Shape Expressions (ShEx) Schemas in Wikidata, see Katherine Thornton et al., "Using Shape Expressions (ShEx) to Share RDF Data Models and to Guide Curation with Rigorous Validation," in *The Semantic Web*, ed. Pascal Hitzler et al., vol. 11503 (Cham: Springer International Publishing, 2019), 606–20, https://doi.org/10.1007/978-3-030-21348-0_39; Katherine Thornton, Kenneth Seals Nutt, and Anne Chen, "Encoding Archaeological Data Models as Wikidata Schemas: Utilizing Shape Expressions to Structure Collaborative Linked Open Data for Digital Storytelling Within the International Dura-Europos Archive," *The International Journal of Technology, Knowledge, and Society* 21, no. 1 (2024): 69–83, https://doi.org/10.18848/1832-3669/CGP/v21i01/69-83.

[11] James Surowiecki, *The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies, and Nations* (New York: Doubleday, 2004).

[12] Yochai Benkler, Aaron Shaw, and Benjamin Mako Hill, "Peer Production: A Form of Collective Intelligence," in *Handbook of Collective Intelligence*, ed. Thomas Malone and Michael Bernstein (Cambridge, Massachusetts: MIT Press, 2015), 175–204; T. Bücheler et al., "Crowdsourcing, Open Innovation and Collective Intelligence in the Scientific Method: A Research Agenda and Operational Framework," in *Artificial Life XII – Twelfth International Conference on the Synthesis and Simulation of Living Systems, Odense, Denmark, 19 August 2010 - 23 August 2010* (Odense: Danemark, 2010), 679–86, https://doi.org/10.5167/UZH-42435.

converge through dialogue and participation. In Wikidata, this manifests through talk pages, project discussions, collaborative schema building, and the continuous negotiation of meanings across languages, cultures, and disciplinary boundaries. It is through this ongoing process that the authority and legitimacy of the platform are constructed—not imposed from above, but co-produced by its users.

Conclusion

We suggest that Wikidata is not merely a technical tool but rather a space where methodological and epistemological debates can unfold. By engaging with this dynamic, researchers can enhance their projects while contributing to the creation of a more sustainable, inclusive, and collaborative knowledge base.

> ! To add
>
> - our presentation invites reflection on the implications of this shift toward distributed authority.
> - How can that shift in authority benefit academic research projects?
> - Is Wikidata's epistemological paradigm coherent with ours?
> - Can we think of a generic epistemological framework to be effectively applied to specific academic endeavors?

References

Benkler, Yochai, Aaron Shaw, and Benjamin Mako Hill. "Peer Production: A Form of Collective Intelligence." In *Handbook of Collective Intelligence*, edited by Thomas Malone and Michael Bernstein, 175–204. Cambridge, Massachusetts: MIT Press, 2015.

Bianchini, Carlo, Stefano Bargioni, and Camillo Carlo Pellizzari Di San Girolamo. "Beyond VIAF: Wikidata as a Complementary Tool for Authority Control in Libraries." *Information Technology and Libraries* 40, no. 2 (2021). https://doi.org/10.6017/ital.v40i2.12959.

Bücheler, T., R. M. Füchslin, R. Pfeifer, and J. H. Sieg. "Crowdsourcing, Open Innovation and Collective Intelligence in the Scientific Method: A Research Agenda and Operational Framework." In *Artificial Life XII – Twelfth International Conference on the Synthesis and Simulation of Living Systems, Odense, Denmark, 19 August 2010 - 23 August 2010*, 679–86. Odense: Danemark, 2010. https://doi.org/10.5167/UZH-42435.

Candela, Gustavo, Mirjam Cuper, Olga Holownia, Nele Gabriëls, Milena Dobreva, and Mahendra Mahey. "A Systematic Review of Wikidata in GLAM Institutions: A Labs Approach." In *Linking Theory and Practice of Digital Libraries*, edited by Apostolos Antonacopoulos, Annika Hinze, Benjamin Piwowarski, Mickaël Coustaty, Giorgio Maria Di Nunzio, Francesco Gelati, and Nicholas Vanderschantz, 34–50. Cham: Springer Nature Switzerland, 2024. https://doi.org/10.1007/978-3-031-72440-4_4.

Cook, Stacey. "The Uses of Wikidata for Galleries, Libraries, Archives and Museums and Its Place in the Digital Humanities." *Comma* 2017, no. 2 (2019): 117–24. https://doi.org/10.3828/comma.2017.2.12.

Fagerving, Alicia. "Wikidata for Authority Control: Sharing Museum Knowledge with the World." *Digital Humanities in the Nordic and Baltic Countries Publications* 5, no. 1 (2023): 222–39. https://doi.org/10.5617/dhnbpub.10665.

Neubert, Joachim. "Wikidata as a Linking Hub for Knowledge Organization Systems? Integrating an Authority Mapping into Wikidata and Learning Lessons for KOS Mappings." In *Proceedings of the 17th European Networked Knowledge Organization Systems Workshop Co-Located with the 21st International Conference on Theory and Practice of Digital Libraries 2017 (TPDL 2017), Thessaloniki, Greece, September 21st, 2017*, edited by Philipp Mayr, Douglas Tudhope, Koraljka Golub, Christian Wartena, and Ernesto William De Luca, 1937:14–25. CEUR Workshop Proceedings. CEUR-WS.org, 2017.

Piscopo, Alessandro, and Elena Simperl. "What We Talk about When We Talk about Wikidata Quality: A Literature Survey." In *Proceedings of the 15th International Symposium on Open Collaboration*, 1–11. Skövde Sweden: ACM, 2019. https://doi.org/10.1145/3306446.3340822.

Severo, Marta. *L'impératif participatif: Institutions culturelles, amateurs et plateformes*. Bry-sur-Marne: Institut National de l'Audiovisuel (INA), 2021. https://doi.org/10.3917/ina.sever.2021.01.

Shenoy, Kartik, Filip Ilievski, Daniel Garijo, Daniel Schwabe, and Pedro Szekely. "A Study of the Quality of Wikidata." *Journal of Web Semantics* 72 (2022): 100679. https://doi.org/10.1016/j.websem.2021.100679.

Surowiecki, James. *The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies, and Nations*. New York: Doubleday, 2004.

Thornton, Katherine, Kenneth Seals Nutt, and Anne Chen. "Encoding Archaeological Data Models as Wikidata Schemas: Utilizing Shape Expressions to Structure Collaborative Linked Open Data for Digital Storytelling Within the International Dura-Europos Archive." *The International Journal of Technology, Knowledge, and Society* 21, no. 1 (2024): 69–83. https://doi.org/10.18848/1832-3669/CGP/v21i01/69-83.

Thornton, Katherine, Harold Solbrig, Gregory S. Stupp, Jose Emilio Labra Gayo, Daniel Mietchen, Eric Prud'hommeaux, and Andra Waagmeester. "Using Shape Expressions (ShEx) to Share RDF Data Models and to Guide Curation with Rigorous Validation." In *The Semantic Web*, edited by Pascal Hitzler, Miriam Fernández, Krzysztof Janowicz, Amrapali Zaveri, Alasdair J. G. Gray, Vanessa Lopez, Armin Haller, and Karl Hammar, 11503:606–20. Cham: Springer International Publishing, 2019. https://doi.org/10.1007/978-3-030-21348-0_39.

Van Veen, Theo. "Wikidata: From 'an' Identifier to 'the' Identifier." *Information Technology and Libraries* 38, no. 2 (2019): 72–81. https://doi.org/10.6017/ital.v38i2.10886.

Zhao, Fudie. "A Systematic Review of Wikidata in Digital Humanities Projects." *Digital Scholarship in the Humanities* 38, no. 2 (2023): 852–74. https://doi.org/10.1093/llc/fqac083.