

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

Mathilde Verstraete Maxime Guénette Marcello Vitali-Rosati

As collaborative digital platforms become more common in cultural and Digital Humanities projects, the role of non-experts and their contributions to these platforms is shifting. Platforms such as Wikidata invite a reconsideration of where authority lies: not solely in the hands of experts, but also within a wider community of contributors who bring diverse forms of expertise. This development challenges hierarchies in academic and cultural institutions and raises new questions about validation, trust and collaborative knowledge production. In this paper, we explore these questions through the lens of the *Anthologia Graeca* project, where the integration of Wikidata has become a way of rethinking editorial authority as a shared process.

Keywords: Greek Anthology, Digital Philology, Authority, Wikidata, Collective Intelligence

Introduction

The management and preservation of research data in the Humanities increasingly raise questions concerning sustainability, accessibility, sharing, and validation. In this context, Wikidata emerges as a powerful and collaborative tool. By challenging traditional models in which researchers act simultaneously as producers and gatekeepers of authority, Wikidata reconfigures these issues and fosters new paradigms of collaborative knowledge production.

Within the framework of Digital Humanities (DH), which emphasizes open processes, data interoperability and collective engagement, Wikidata functions as a knowledge base, enabling a sort of collective verification and semantic linking of information. Unlike traditional academic publishing, where authority is centralised and often restricted to established institutions or recognised experts, Wikidata operates through a model of continuous, multilingual, and community-based editing that promotes the dissemination of free and accessible knowledge globally. This paradigmatic shift invites a fundamental rethinking of authority, editorial responsibility, and the epistemological foundations of data.

How, then, can *expert-led* projects — whether developed by academics, government agencies or GLAM institutions (Galleries, Libraries, Archives and Museums) — work with a generalist platform such as Wikidata to generate new forms of knowledge? How do these hybrid models, which combine scholarly expertise with public participation, challenge traditional boundaries between academic and amateur contributors, and between knowledge production and validation?

In this article, we explore how the infrastructure and logic of Wikidata can be integrated into DH projects, focusing on the *Anthologia Graeca* (AG) project, a collaborative digital edition of the *Greek Anthology*. After outlining the role of Wikidata in DH initiatives, we turn to our case study to analyse how Wikidata is embedded in the data model of the AG project and contributes to the emergence of new spaces of knowledge creation. Finally, we reflect on the role of authority and collective intelligence in shaping such projects and reflections.

Wikidata and Digital Humanities

Since its creation in 2012, Wikidata has gradually become one of the most significant knowledge graphs on the Web. As structured data plays an increasingly central role in the organization, retrieval, and circulation of information online, Wikidata occupies a pivotal position in shaping how knowledge is represented, accessed, and reused.

As a result, Wikidata has gained popularity within the field of DH. Although there has been — and to some extent, continues to be — skepticism about its quality and potential as a scholarly resource, recent studies have shown that it is now widely adopted across DH projects (Cook 2019; Zhao 2023). Within this context, it is primarily used as a *content provider* — a means to access, publish, or disseminate Linked Open Data (LOD) while avoiding many of the technical and financial constraints traditionally associated with the Semantic Web.

For institutions in the GLAM sector (Galleries, Libraries, Archives, and Museums), Wikidata serves a somewhat different role: as a *publishing platform* and a *tool for metadata curation*. Because digitization strategies have historically been developed independently from one institution to another, accessibility and discoverability vary widely in the GLAM domain (Fagervang 2023). Wikidata is thus used either to publish digital identifiers for cultural heritage objects—enriched with LOD for the first time—or to enhance metadata by linking institutional records to Wikidata, improving visibility and interoperability (Candela et al. 2024).

Wikidata also plays a *connective role*. Its structure as a knowledge graph and its alignment with LOD principles allow it to act as a bridge between otherwise siloed datasets. By assigning persistent identifiers and promoting alignment across vocabularies and standards, Wikidata enables interoperability between different projects and institutions.

Wikidata as a linking hub

One of the key benefits of Wikidata, as identified by many DH projects and GLAM institutions, is in its function as a hub for linking heterogeneous external resources and datasets (Neubert 2017). This is achieved through the creation of external *identifier properties*, which connect entities across different databases to their corresponding Wikidata items.

Take, for example, the case of Megara, an ancient Greek *(pólis)* located in the northern part of the Isthmus of Corinth. Its Wikidata item ([Q42307600](#)) is linked to a wide range of sources, including dictionaries, library catalogues and domain-specific databases on Greco-Roman antiquity, such as [Pleiades](#), [ToposText](#) and [MANTO](#). Through such links, Wikidata facilitates seamless cross-referencing and acts as a form of authority control (Fagervig 2023). As LOD is designed to enrich data with contextual relationships, it becomes much more efficient to enrich a dataset either directly through Wikidata or through cross-queries across databases all linked to a common Q-item.

Some scholars have gone further, suggesting that Wikidata should not only serve as a hub for linking external identifiers, but should actually function as *the* reference identifier (Van Veen 2019). The multiplication of identifiers for the same entity (such as [Megara](#)) can lead to confusion and poor data reconciliation over time. Adopting Wikidata as a universal identifier could offer several benefits: a unified description model, a single SPARQL endpoint and API for querying, and a sustainable infrastructure for data access and storage.

While this position may seem more radical, it highlights another important aspect of Wikidata's relationship with external resources: that of reciprocal contribution. As demonstrated in a paper on the interaction between Wikidata and VIAF — a global platform that aggregates name authority files from multiple institutions — bi-directional comparison and contribution can improve data quality on both sides (Bianchini, Bargioni, and Pellizzari Di San Girolamo 2021). In this way, GLAM institutions such as libraries can not only contribute to Wikidata, but also benefit from it, strengthening a collaborative ecosystem of knowledge validation and enrichment.

Authority on Digital Platforms

This brings us to a central question: how can academic projects and Wikidata benefit from one another when unequal hierarchies of authority persist between these platforms? Indeed, there is a clear distinction between contributions from experts — scholars, government bodies, or institutions — and those from non-experts, such as general Wikidata editors. In academia, authority is frequently equated with expertise and influence, recognized and exerted within the boundaries of specific disciplines. Outside academia, institutions such as governments or GLAM organizations are similarly perceived as authoritative sources. Within any scholarly domain, certain individuals — by means of influential publications or institutional standing — are considered the custodians of knowledge. This interpretation of the authoritative figure

typically establishes a top-down hierarchy, privileging traditional producers of knowledge — positioning them as its sole custodians.

We argue that this top-down hierarchy is contextual and is not always what is best for knowledge dissemination, especially for digital platforms. As collaborative, crowdsourced, or 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 data’s quality.

Thus, the status of Wikidata’s editors as an authority figure is questioned because they are mostly seen by scholars and professionals as “amateurs”.¹ 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 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 (Piscopo and Simperl 2019; Shenoy et al. 2022; Zhao 2023). Its growing community actively contributes to its enrichment, 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 ((on Shape Expressions (ShEx) Schemas in Wikidata, see Thornton et al. 2019; Thornton, Seals Nutt, and Chen 2024)). 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.

¹We 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 Severo (2021)

Another way to reconceptualize and test the notion of authority in DH projects is to fully integrate Wikidata at the core of their infrastructures—not simply as an external reference point, but as a foundational layer of data modeling, curation, and publication. In doing so, the authoritative role traditionally held by academic experts is shared with Wikidata editors. This shift reframes authority not as a fixed attribute of disciplinary recognition, but as an emergent property of collaborative knowledge practices. It also requires researchers to reconsider their position—not above or outside the platform, but alongside a distributed network of contributors who co-construct meaning, value, and trust in data.

In order to better understand these issues, we turn to our case-study, the AG project. We will explore how the AG project is building its digital platform through collaborative participation, collective intelligence, and Wikidata to question the status of authoritative figure in academia.

Context

Since 2014, Marcello Vitali-Rosati and his team have been developing a digital and collaborative edition of the *Greek Anthology* (Verstraete and Mellet 2024; Verstraete 2024; Vitali-Rosati et al. 2021; Mellet 2020; Vitali-Rosati et al. 2020). The project was born out of the need to index and make accessible this foundational corpus, which gathers nearly all the known epigrams of ancient Greek literature. This project has led to [the creation of a platform](#). Drawing on the contributions of a wide range of collaborators, it provides, for each epigram a cluster of information. Each epigram, thus, has its own page where one can find (1) its position in the *Palatine* manuscript (the *codex Palatinus graecus* 23, the principal testimony for the *Palatine Anthology*), retrieved via the IIIF protocol through the Heidelberg Library’s annotation tool linked to its API, (2) multiple translations in various languages, (3) various keywords (author, cities and other thematic keywords), (3) commentaries, (4) internal and external references, and (5) cross-alignments between translations. [A REST API is available](#) for querying the dataset.

This case-study focuses specifically on the use of keywords—covering entities such as (a) authors, (b) cities, and (c) others keywords collections divided in sections like deities, epithets, and epiclesis – [the full list can be found on the website](#). As part of the project’s development, we introduced a rule requiring all keywords to be linked to Wikidata: any new keyword must be associated with a corresponding Wikidata identifier. This policy prompted a comprehensive reconciliation of our existing metadata with the Wikidata knowledge base.

History of the platform(s)

Since 2014, our team has been developing a digital editorial model tailored to the *Greek Anthology*, an ancient corpus, similar to a fragmentary one. Early stages of the project focused on exploring how digital tools could enhance both access to and understanding of that material.

Initial prototypes—such as a SPIP-based website—provided a foundation for collaborative enrichment and allowed us to begin identifying the technical and hermeneutic challenges of such an edition.

These early experiments revealed that designing a digital edition involves more than providing access (Sahle (2016)): it requires a coherent epistemological model to structure the relationships among texts, editions, translations, annotations, and contributors. Over time, the platform evolved to support collaborative editing, allowing users to contribute translations, metadata, and commentary: that is when the platform *Anthologia Palatina* was created. Yet, as the project expanded and new collaborators joined, the limits of the initial technical infrastructure became increasingly apparent—particularly regarding multilingualism and the lack of stable identifiers for core entities like authors, cities, or any kind of tag represented by a keyword.

Building on an initial REST API, a second platform was developed – *Anthologia graeca, augmented with an API*. It uses GraphQL and the backend schema is implemented using Django, a Python-based Web framework. We focused on collaborations (with Perseus, Perseids, the Library of Heidelberg, by example), highlighting the importance of strong data structuring and interoperability, core principles in the Digital Classics community. Our model had to be compatible with broader efforts in open scholarship and heritage valorization. The platform introduced a more granular editorial model centered on the epigram as a fundamental unit. This new platform adopts a semantic Web approach and is grounded in the systematic use of Wikidata identifiers², marking a significant evolution in our editorial framework. This decision is a turning point. Not only does it resolve earlier issues around entity identification, but it also reaffirms the project’s original commitment to collaborative editorial practices: by aligning with Wikidata, we ensure that each new contribution is embedded in a broader, federated knowledge network.

The Authors of the *Greek Anthology*

Wikidata thus plays a central role in the AG project. From the new platform on, all new keywords used to annotate the platform are created via a Wikidata URI – note that some old keywords still have not been linked to a Wikidata ID yet, as we said above. This integration

²Note that all data from previous platforms have automatically been imported into this one, though not all entities are currently linked to Wikidata—an ongoing task. This transition reflects our evolving editorial model: one that is increasingly open, decentralized, and connected to the wider ecosystem of digital knowledge production. Another second note is that the project is not finished — and never will be. This is not just a matter of ongoing development; it’s inherent to the nature of the corpus itself. It’s an anthology — a form that, by definition, resists closure. New readings, translations, annotations, and connections will always be possible. At present, nearly all the epigrams are available on the platform, but not all of them. Some are still being reviewed, others are pending transcription or metadata enrichment. And beyond the texts themselves, the work of annotation, translation, and linking — especially through tools like Wikidata — opens up an infinite horizon of contributions. The platform is designed to remain open, both structurally and epistemologically, to future layers of meaning and interpretation.

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 DO

- Chantier Auteurs: retour d'expérience? méthodo? stat(-ish)?
- Interactions avec la communauté

Quelques problématiques

On co-existing informations and world visions

! 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

! To add (?)

- à relier aux auteurs surtout ?

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?

Collective Intelligence

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 (Surowiecki 2004). 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 (Benkler, Shaw, and Mako Hill 2015; Bücheler et al. 2010).

As Pierre Lévy argues, collective intelligence is especially interesting and valuable for digital environments (Lévy 1994). Collaborative digital platforms are not about homogenizing knowledge or erasing differences in expertise, but about creating a space where knowledge can circulate, evolve, and converge through dialogue and participation. In Wikidata, this manifests through talk pages, project discussions, creation or modification of items and properties, 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. 2015. “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.
- Bianchini, Carlo, Stefano Bargioni, and Camillo Carlo Pellizzari Di San Girolamo. 2021. “Beyond VIAF: Wikidata as a Complementary Tool for Authority Control in Libraries.” *Information Technology and Libraries* 40 (2). <https://doi.org/10.6017/ital.v40i2.12959>.
- Bücheler, T., R. M. Füchslin, R. Pfeifer, and J. H. Sieg. 2010. “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: Denmark. <https://doi.org/10.5167/UZH-42435>.
- Candela, Gustavo, Mirjam Cuper, Olga Holownia, Nele Gabriëls, Milena Dobрева, and Mahendra Mahey. 2024. “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. https://doi.org/10.1007/978-3-031-72440-4_4.
- Cook, Stacey. 2019. “The Uses of Wikidata for Galleries, Libraries, Archives and Museums and Its Place in the Digital Humanities.” *Comma* 2017 (2): 117–24. <https://doi.org/10.3828/comma.2017.2.12>.
- Fagerving, Alicia. 2023. “Wikidata for Authority Control: Sharing Museum Knowledge with the World.” *Digital Humanities in the Nordic and Baltic Countries Publications* 5 (1): 222–39. <https://doi.org/10.5617/dhnbpub.10665>.
- Lévy, Pierre. 1994. *L’intelligence collective : pour une anthropologie du cyberspace*. Paris: La Découverte.
- Mellet, Margot. 2020. “Penser Le Palimpseste Numérique. Le Projet d’édition Numérique Collaborative de l’anthologie Palatine.” *Captures* 5 (1). <https://doi.org/10.7202/1073479ar>.
- Neubert, Joachim. 2017. “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.
- Piscopo, Alessandro, and Elena Simperl. 2019. “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. <https://doi.org/10.1145/3306446.3340822>.
- Sahle, Patrick. 2016. “2. What Is a Scholarly Digital Edition?” In *Digital Scholarly Editing : Theories and Practices*, edited by Matthew James Driscoll and Elena Pierazzo, 19–39. Digital Humanities Series. Cambridge: Open Book Publishers. <https://books.openedition.org/obp/3397>.
- Severo, Marta. 2021. *L’impératif participatif: Institutions culturelles, amateurs et plateformes*. Bry-sur-Marne: Institut National de l’Audiovisuel (INA). <https://doi.org/10.3917/ina.sever.2021.01>.
- Shenoy, Kartik, Filip Ilievski, Daniel Garijo, Daniel Schwabe, and Pedro Szekely. 2022. “A Study of the Quality of Wikidata.” *Journal of Web Semantics* 72:100679. <https://doi.org/10.1016/j.websem.2021.100679>.
- Surowiecki, James. 2004. *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.
- Thornton, Katherine, Kenneth Seals Nutt, and Anne Chen. 2024. “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 (1): 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 Mitichen, Eric Prud’hommeaux, and Andra Waagmeester. 2019. “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. https://doi.org/10.1007/978-3-030-21348-0_39.
- Van Veen, Theo. 2019. “Wikidata: From ‘an’ Identifier to ‘the’ Identifier.” *Information Technology and Libraries* 38 (2): 72–81. <https://doi.org/10.6017/ital.v38i2.10886>.
- Verstraete, Mathilde. 2024. “Dynamiques Actoriales Au Sein Du Projet d’édition Numérique Collaborative de l’ Anthologie Grecque.” *Sens Public*. <http://sens-public.org/articles/1692/>.
- Verstraete, Mathilde, and Margot Mellet. 2024. “Passés et présents anthologiques. Le projet d’édition numérique collaborative de l’Anthologie grecque.” In *Communautés et pratiques d’écritures des patrimoines et des mémoires*, Presses universitaires de Paris-Nanterre. Paris.
- Vitali-Rosati, Marcello, Margot Mellet, Servanne Monjour, Antoine Fauchié, Timothée

- Guicherd, David Larlet, and Enrico Agostini-Marchese. 2021. “L’épopée numérique de l’Anthologie grecque : entre questions épistémologiques, modèles techniques et dynamiques collaboratives.” *Sens Public*, July.
- Vitali-Rosati, Marcello, Servanne Monjour, Joana Casenave, Elsa Bouchard, and Margot Mellet. 2020. “Editorializing the Greek Anthology: The Palatin Manuscript as a Collective Imaginary.” *Digital Humanities Quarterly* 014 (1).
- Zhao, Fudie. 2023. “A Systematic Review of Wikidata in Digital Humanities Projects.” *Digital Scholarship in the Humanities* 38 (2): 852–74. <https://doi.org/10.1093/llc/fqac083>.