

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 (Fagervig 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* that privileges traditional producers of knowledge — positioning them as its sole custodians.

We argue, however, that this top-down hierarchy model is contextual and not always best suited for knowledge dissemination, particularly in the case of digital platforms. As collaborative, crowdsourced, and peer-contributed infrastructures like Wikidata increasingly assert themselves as alternative forms of authority — both within and beyond academia — they challenge conventional assumptions about data reliability and the guaranty of expertise. This is mainly due to the fact that, in most cases, anyone can contribute to these platforms, leading some to question the quality of the data.

The status of Wikidata editors as authoritative actors is therefore often questioned. Many scholars and professionals consider them to be *amateurs* rather than experts.¹ Despite the increasing involvement of non-professionals in cultural initiatives, especially online, their contributions often remain undervalued or require validation by recognised professionals to be considered reliable. This tension highlights the ambiguous relationship between new actors in knowledge creation and traditional epistemic authorities who are still unsure about how to integrate them.

On Wikidata, this dynamic manifests in conflicting perceptions of authority. While the platform is designed to support collaborative knowledge production and open participation, its data is frequently considered trustworthy only when curated or approved by recognized experts. Yet, recent studies suggest that although Wikidata still has room for improvement, the platform is increasingly being recognised as a high-quality knowledge graph — one whose quality depends on context and must be assessed on a case-by-case basis (Piscopo and Simperl 2019; Shenoy et al. 2022; Zhao 2023). The platform’s growing and active community contributes significantly to its data quality, while tools such as Shape Expressions (ShEx) Schemas are being implemented to enforce model conformity and internal consistency ((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.

One of the most effective ways to ensure the quality and relevance of Wikidata is not only to assess it from the outside, but rather to engage directly with it. Through engagement with the platform — making statements, creating data models, correcting mistakes, and arguing over ontologies — researchers, cultural institutions and engaged amateurs alike help to shape the platform and its knowledge graph. Wikidata is not a finished product to be critiqued, but an open, iterative space where quality emerges through collective interaction.

¹We define *amateur* as a socially engaged, non-professional actor who contributes to cultural, artistic or documentary activities outside formal institutional structures - often through participatory digital platforms. This figure is typically characterised by a high degree of self-taught expertise, an individual pursuit of excellence, a paradoxical sociability rooted in both emancipation and community, and a privileged relationship with digital platforms as new spaces of recognition and knowledge production. On the evolution of the term, see Severo (2021)

Another way to reconceptualise authority in digital humanities projects is to integrate Wikidata not only as a reference point, but as a core infrastructural layer for data modelling, curation and publication. In this way, the authoritative role traditionally held by academic experts is distributed and shared with a broader community of Wikidata contributors. Authority is thus reframed — not as a fixed attribute derived from institutional status, but as an emergent property of collaborative practices. Such a reframing invites researchers to reconsider their own position — not above or outside the platform, but alongside a distributed network of contributors who collectively construct meaning, value, and trust in data.

To explore these dynamics in more depth, we now turn to our case study: the *AG* project. We will examine how the project builds its digital infrastructure around Wikidata, collective intelligence and participatory practices to challenge and reshape the status of authoritative figures within academic knowledge production.

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 render accessible this foundational corpus, which gathers nearly all the known epigrams of ancient Greek literature. This project has led to the creation of [the platform *anthologiagraeca*](#), which, through the contributions of a wide range of collaborators, offers a rich set of information for each epigram. Each epigram is presented on a dedicated page, where one can access (1) its location in the *Palatine* manuscript (the *codex Palatinus graecus* 23, the principal testimony for the *Palatine Anthology*), retrieved via the IIF protocol through the Heidelberg Library’s annotation tool linked to its API², (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](#). In the course of the platform’s development, a rule was introduced requiring that each keyword be linked to Wikidata: any new keyword must be associated with a corresponding Wikidata identifier. This editorial choice initiated a comprehensive reconciliation of the platform’s metadata with the Wikidata knowledge base.

²See [Heidelberg, Universitätsbibliothek, *Pal. gr.* 23](#). The final folios of the manuscript (615-709) are held at the BNF, under the shelfmark [Paris. Suppl. gr. 384](#) ; they are not yet integrated into the platform.

History of the platform(s)

Since 2014, our team has aimed to develop a digital editorial model suited to the *Greek Anthology*, an ancient corpus that resembles a fragmentary tradition in its structure and transmission. 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 to texts (Sahle 2016): it requires a coherent epistemological model to organize the relationships among texts, editions, translations, annotations, and contributors. As the project progressed, 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 corpus expanded and new contributors joined, the limits of the initial infrastructure—particularly in terms of multilingual support and stable identification of entities—became increasingly apparent.

In response, a second platform was developed: *Anthologia Graeca*, augmented by an API, and based on a backend architecture implemented in Django, a Python-based web framework. This iteration emphasized interoperability and collaboration, aligning with core principles of the Digital Classics community. Partnerships—with initiatives such as Perseus, Perseids, the Heidelberg Library, or the *Liceo Classico Cagnazzi* (Altamura, Bari) — highlight the project’s orientation toward open scholarship and cultural heritage valorization. The editorial model was also refined: the epigram became the fundamental editorial unit, and the platform adopted a semantic web approach grounded in the systematic use of Wikidata identifiers³. The adoption of Wikidata thus marks a significant turn in the project’s editorial framework. It addresses long-standing challenges around entity disambiguation and contributes to embedding each editorial act within a federated, multilingual knowledge network. This choice also reaffirms the project’s original ambition: to develop a truly collaborative and open-ended editorial model.

The Authors of the *Greek Anthology*

Wikidata thus plays a central role in the *AG* project. From the new platform onwards, all new keywords used to annotate the epigrams — including authors — are systematically created via

³All data from previous platforms were automatically imported, although not all entities are yet linked to Wikidata — this reconciliation is ongoing, but reveals some epistemological problems (see *infra*). Importantly, the project is unfinished—and necessarily so. This is not simply due to the usual incompleteness of digital infrastructure, but because of the anthology’s very nature: a form that resists closure. New readings, translations, annotations, and interconnections will always be possible. While nearly all epigrams are already available online, some remain unfinished, and others await transcription or metadata enrichment. Annotation, translation, and semantic linking — particularly through Wikidata — open up a boundless field of contribution. The platform is designed to remain open, both structurally and epistemologically, to future layers of meaning and interpretation.

a Wikidata URI. Some earlier entries have not yet been reconciled, but the ongoing process of linking each entity to a stable identifier has already significantly improved the platform's consistency and interoperability. This integration proved especially valuable in addressing inconsistencies within our list of authors. Because both Wikidata and our platform's data model support multilingual data, discrepancies such as missing names, duplicate entries, and variations across languages (French, English, Italian, Ancient Greek, and Latin) became visible at the data level (see [the list of authors][https://anthologiagraeca.org/authors/\(\)](https://anthologiagraeca.org/authors/)). To resolve these issues, we first ensured that every author in our database was associated with a Wikidata URI. We then enriched Wikidata itself by adding missing multilingual labels and aliases. Once this information was uploaded, the Wikidata community rapidly reviewed and normalized it according to their standards. This process improved not only the reliability of our own metadata, but also the quality of Wikidata as a federated knowledge base.

! 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 ?

Conclusion

Wikidata is not just a technical tool for data storage and retrieval; it is a dynamic epistemic space where academic knowledge is collaboratively shaped, revised, and legitimized. By delegating (while taking part of) some of our curatorial authority to Wikidata, the *AG* project participates in a broader shift towards distributed models of knowledge production (Benkler, Shaw, and Mako Hill 2015; Bücheler et al. 2010). This delegation does not imply an abandonment of academic rigor, but rather a reconfiguration of expertise — one that recognizes the value of collective intelligence (Lévy 1994) and the potential of community-driven platforms to maintain high-quality, multilingual, and interoperable data (Surowiecki 2004).

The implications of this shift are manifold. It invites researchers to rethink their role not only as producers of content, but also as facilitators of open, dialogical knowledge systems. It also raises critical questions: To what extent can academic standards be reconciled with the epistemological norms of a platform like Wikidata? How do we balance openness with accuracy, and participation with control? Far from erasing these tensions, the integration of Wikidata into our project foregrounds them — and in doing so, creates an opportunity for scholars to collectively reflect on the processes of validation, authorship, and authority.

Ultimately, we suggest that embracing platforms like Wikidata means accepting that academic knowledge is always provisional, situated, and co-produced. It is a move toward a more resilient and plural form of scholarship - one that recognizes the potential of the many without losing sight of the responsibility of the few.

! 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. Fuchslin, 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/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 Auctoriales 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/lhc/fqac083>.