

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

Maxime Guénette

Mathilde Verstraete

May 6, 2025

Title page @max

Hello, and thank you all for being here today. We're excited to share with you some key insights from our recent work at the intersection of Wikidata and Greek philology, particularly within the framework of a collaborative digital edition of the *Greek Anthology*.

Introduction @Max @Math

The management and preservation of research data in the Humanities increasingly raise questions concerning their sustainability, accessibility, sharing, and validation.

In this context, Wikidata stands out as a powerful and collaborative platform. By challenging traditional models in which researchers act at the same time 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 often emphasises 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 us to rethink authority, editorial responsibility, and even the epistemological foundations of data.

How, then, can expert-led projects — whether developed by academics, government agencies or GLAM institutions — work with a 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?

Overview @Max

In this talk, we explore how the infrastructure and logic of Wikidata can be integrated into DH projects.

Our focus will be on *Anthologia Graeca* — a collaborative digital edition of the *Greek Anthology* (Slide).

We'll start by situating Wikidata within the broader landscape of DH initiatives (Slide). Then we'll turn to *Anthologia Graeca* as a case study, and look more closely at how Wikidata is embedded in the project's data model — particularly through our treatment of the keyword “authors” — and how this integration opens new spaces for knowledge production (Slide).

Finally, drawing on those concrete examples, we will reflect on the tensions between authority and collective intelligence that shape such initiatives, and the ways in which digital infrastructures can both challenge and extend traditional scholarly practices.

Wikidata and Digital Humanities @Max

Since it was launched in 2012, Wikidata has emerged as one of the most important knowledge graphs on the Web. With increased importance for data structuring and sharing, Wikidata holds a central place in accessing and reusing knowledge.

Despite initial skepticism, it is now widely adopted in DH as a flexible platform for publishing and using Linked Open Data (LOD) without needing deep technical skills of Semantic Web standards. In the GLAM sector, so Galleries, Libraries, Archives, and Museums, Wikidata is often used to curate metadata and publish cultural heritage records to increase discoverability and interoperability.

Wikidata as A Linking Hub @Max

One of Wikidata's major strengths, widely recognized the digital community, lies in its role as a central hub for linking heterogeneous datasets. This is made possible through external identifier properties, which connect Wikidata items to corresponding entities across databases.

Take for example Megara, an ancient Greek city near Corinth. Its Wikidata item links to a variety of sources — from library catalogues to specialized databases like MANTO (Slide),

Mythoskop (**Slide**), and Pleiades (**Slide**). These links enable cross-referencing, facilitate authority control, and support the enrichment of data through Linked Open Data (LOD). By relying on Wikidata’s Q-items, researchers can efficiently integrate and query related resources across platforms.

Some scholars argue that we should go even further and make Wikidata the unique reference identifier. Multiple identifiers for the same entity — such as Megara — risk fragmenting data and complicating reconciliation. Using Wikidata as a single ID could offer a unified data model, a single SPARQL endpoint, and a long-term infrastructure perrenity for querying and storage.

Although this position is more ambitious, it underlines an essential aspect of Wikidata’s ecosystem: reciprocal contribution. DH projects and GLAM institutions can both feed into and benefit from Wikidata, ultimately improving the quality of each others data.

Authority on Digital Platforms @Max

This brings us to a central question: how then can academic projects and Wikidata mutually enrich one another, given their different systems of validation and authority? While many Wikidata editors have strong expertise, the platform is built on openness, collaborative editing, and peer consensus — rather than formal academic credentials. By contrast, academia tends to associate authority with institutional affiliation, scholarly output, and disciplinary recognition. Similarly, GLAM institutions or government agencies are seen as authoritative due to their structured oversight and perceived reliability.

These differing frameworks create asymmetries in how contributions are trusted. Academia often relies on top-down models of validation, while Wikidata operates through distributed consensus. Understanding these dynamics is essential for building stronger collaboration between both ecosystems.

We argue that traditional hierarchies are not always the most effective model for digital knowledge production. Collaborative platforms such as Wikidata challenge long-standing assumptions about authority by enabling anyone to contribute. This openness raises concerns about data quality, and Wikidata editors are often viewed as amateurs or “citizen scientists.” Their work is frequently only legitimised when endorsed by academic or institutional figures.

Despite these tensions, Wikidata is increasingly recognised as a high-quality knowledge graph. Its reliability is context-dependent and must be assessed case by case. Its community plays a key role in maintaining quality, supported by tools like Shape Expressions (ShEx) that enforce data model consistency. These community-driven mechanisms represent a decentralised but structured approach to data governance.

Crucially, improving Wikidata is not just a matter of external assessment but of active engagement. Researchers, GLAM professionals, and contributors from all backgrounds can help

shape the platform by adding statements, correcting errors, refining data models, and discussing ontologies. Wikidata is not a finished product but a dynamic infrastructure where quality emerges through interaction.

To rethink authority in DH, we suggest using Wikidata not just as a reference source but as a foundational layer for data modelling, curation, and publication. In this model, epistemic authority becomes a shared process — emerging from collaboration rather than imposed from above. Academic researchers contribute alongside volunteers, curators, and technologists in shaping meaning, structure, and trust within the data.

To explore this further, we now turn to our case study: the *Anthologia Graeca* 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.

A collaborative and digital edition of the *Greek Anthology* @Math

The case study we'll be focusing on today — and from which our reflections emerge — is the *Anthologia Graeca* project, a collaborative digital edition of the *Greek Anthology*.

This project is led by Marcello Vitali-Rosati as part of his Canada Research Chair in Digital Textualities, in collaboration with co-researchers Christian Raschle and Elsa Bouchard. I am the coordinator of the project since 2021.

The project has been supported by multiple grants from the Social Sciences and Humanities Research Council of Canada.

It brings together a large and diverse team: scholars, developers, editors, students from both the University of Montreal and the University of Naples, as well as even high school students from Bari, Italy. Institutional partners include the CRIHN (Interuniversity Research Centre on the Digital Humanities), the GREN (Research Group on Scholarly Editions in a Digital Context), Perseus and Perseids, Heidelberg University Library, and many more.

This rich and distributed collaboration provides a perfect context for thinking through how Wikidata can reshape editorial models and the distribution of authority in scholarly projects.

Project's corpus @Math

A few words about the corpus so we know what we're talking about.

The *Greek Anthology* is a collection of ancient Greek epigrammatic poetry, spanning over 15 centuries — from the Classical period to the Byzantine era.

At its core is the *Palatine Anthology*, preserved in a 10th-century manuscript held in Heidelberg — the codex Palatinus graecus 23. This is traditionally considered the main source.

To this we add the *Appendix Planudea*, a 13th-century compilation based on different sources. Together, they form what we refer to as the *Greek Anthology* — a corpus that we like to define as open, open both in scope and in structure.

According to our API, it currently includes 4,134 epigrams attributed to 311 authors.

Initially, our platforms focused only on the *Palatine Anthology*. But in a second phase, we extended the scope to cover the broader *Greek Anthology*, in line with our commitment to a more evolving, open, and inclusive corpus.

Project's goals @Math

The project began, somewhat unofficially, in 2014, when Marcello Vitali-Rosati went looking for an epigram he remembered from high school — but couldn't find easily. Quite a few resources were available on the Web: the manuscript was already digitized, the Greek text was online, and commentaries existed — but everything was scattered, with no consistent indexing.

That's where the idea for *Anthologia Graeca* came from: to create a central hub that brings together manuscripts, texts, translations, keywords linked to Wikidata, commentaries, and references — all in an open and interoperable format.

Today, the platform is built on Django, with a public API that lets anyone retrieve the data. It's still a work in progress: some content is missing and need to be added, some features are still in development — like a better referencing to editions used. But all in all, the goal remains clear — to make this corpus accessible, extensible, and richly interconnected.

A few words on the previous platforms @Math

From the beginning, the goal wasn't just to give access to the texts, but to experiment with a digital editorial model suited to the *Greek Anthology*'s fragmentary nature and complex transmission.

We started with a first prototype using SPIP, which helped us explore the indexing of our corpus. This led to a second platform: *Anthologia Palatina*, focused on the *Palatine Anthology* only, and allowing users to contribute translations, metadata, and commentary.

As the corpus expanded and new contributors joined, we faced increasing challenges — especially around multilingualism and entity identification. That's when we moved to the current platform: *Anthologia Graeca*, with a new data model, multilingual support, and a public API.

The platform's keywords @Math

Some comments on the outcome and future work @Math

Conclusion @Math

In conclusion, Wikidata is not just a technical tool — it is a collaborative epistemic space where academic knowledge is shaped, revised, and legitimised. In the AG project, we delegate part of our curatorial authority to Wikidata, embracing a shift toward distributed models of knowledge production. This does not weaken academic rigour but redefines expertise through collective intelligence and the potential of community-driven platforms to support high-quality, multilingual, interoperable data.

This shift challenges researchers to act not only as content producers but also as facilitators of open knowledge systems. It raises key questions: Can academic standards align with Wikidata's norms? How do we balance openness with accuracy?

Instead of smoothing over existing tensions, integrating Wikidata brings them to the surface, prompting deeper questions about how knowledge is validated, attributed, and governed. Engaging with such a platform means accepting that scholarship is shaped through collaboration and context, and that it rarely rests on fixed foundations. Thank you!