

# 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 [30sec]

Hello, and thank you all for being here today. We're excited to share with you some key insights from our recent work on Greek philology and Wikidata.

## Introduction @Max [1mn30sec]

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 Digital Humanities (DH), where openness, interoperability, and collective engagement are often central, Wikidata serves as a knowledge base that enables collective verification and semantic linking. 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 — 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, combining scholarly expertise with public participation, challenge

traditional boundaries between academic and *amateur* contributors, and between knowledge production and validation?

## **Overview @Max [45sec]**

In this talk, we explore how the infrastructure and logic of Wikidata can be integrated into DH projects (**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 [45sec]**

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 [1min30sec]**

One of Wikidata's major strengths, widely by 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 Megara, the 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 connections support cross-referencing, facilitate authority control, and enrich data through Linked Open Data (LOD). By relying on Wikidata's Q-items, researchers can efficiently query and integrate related resources across platforms.

Some scholars argue that we should go further and adopt Wikidata as the unique reference identifier. Multiple identifiers for the same entity — like we’ve seen with 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 sustainability for data querying and storage.

While this approach is more ambitious, it highlights an essential feature of Wikidata’s ecosystem: reciprocal contribution. DH projects and GLAM institutions can both feed into and benefit from Wikidata, ultimately enhancing the quality of each other’s data.

## **Authority on Digital Platforms @Max [3min]**

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.

Wikidata is not a finished product but a dynamic infrastructure where quality emerges through interaction. Researchers, GLAM professionals, and contributors from all backgrounds can help shape the platform by adding statements, correcting errors, refining data models, and discussing ontologies.

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.

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 [45sec]**

The *Greek Anthology* is a vast and fragmentary collection of epigrammatic poetry – mainly, short poems. It comprises more than 4.000 epigrams, written by over 300 hundred authors spanning from the Classical period to the Byzantine era — nearly fifteen centuries of literature. The *Anthology* is the result of ongoing compilation, transmission, and reorganization.

Our aim was to design a digital platform that could reflect this dynamic, evolving structure — one that embraces the *Anthology*'s fluidity instead of imposing artificial boundaries.

## **Project Origins and Team [1min15sec]**

The project was launched in 2014 by Marcello Vitali-Rosati, as part of the Canada Research Chair in Digital Textualities. I've been the project coordinator since 2021.

Supported by several grants from the Social Sciences and Humanities Research Council of Canada, it brings together a wide network of scholars, developers, and students — from the University of Montreal, the University of Naples, and even high school students from Bari. We have many institutional partners, for example the Interuniversity Research Centre on the Digital Humanities, Perseus, or the Heidelberg University Library.

From the beginning, the project has been grounded in the principles of open science: sharing data, working collaboratively across institutions and disciplines, and building tools that remain accessible and reusable. In that spirit, we turned early on to Wikidata — not only as a resource we could draw from, but as a platform we could also contribute to. This engagement has allowed us to rethink how authority and knowledge circulation function in collaborative digital scholarship.

## **Platform and Goals @Math [1min30sec]**

Our goal was to build a digital infrastructure that acts as a central hub — bringing together information and data related to each epigram: main manuscript images, transcriptions, translations, keywords, commentaries, and references. All of this is accessible in an open and interoperable platform to support reuse, collaboration, and long-term accessibility.

The current platform runs on Django, with a public API. While the platform is still evolving — with ongoing work on metadata, referencing, and editorial enrichment — its purpose remains clear: to support collaborative scholarship, semantic annotation, and multilingual engagement with our corpus.

One important aspect I want to highlight is that the project didn't start with a fixed editorial model — it evolved over time, across several platforms. As both our technical and philological understanding matured, so did our goals. This evolution played a crucial role in the reconciliation work I'll describe in a moment.

For instance, in our current platform, adding a new keyword now requires linking it to a corresponding Wikidata identifier. But much of the data that we imported from previous versions — built before we implemented this rule — lacks those identifiers. This mismatch between imported data and current editorial standards is precisely what made reconciliation necessary.

## **The platform's keywords @Math [45sec]**

Our platform uses keywords to describe and organize the content of each epigram. These keywords fall into several categories:

1. Authors to whom the epigrams are attributed ;
2. Cities;
3. And a broad set of general keywords, which include things like collections, deities, cited poets, famous historical or mythical figures, metrical forms, etc.

The first step of our reconciliation work focused on the authors — both because of their centrality in the corpus and because they provided a manageable entry point for aligning our data with Wikidata.

I'll start by walking you through that process, and then briefly say a few words on the other keywords.

## **Reconciling Keywords with Wikidata @Math [1min15sec]**

During the process of reconciling our authors' names with their Wikidata identifiers, we took the opportunity to check the attributions that had been made on our platform and to add the authors' names in different languages to Wikidata.

To do so, we began by exporting the authors into a spreadsheet. That gave us a first overview of inconsistencies — duplicates, spelling variants, ambiguous attributions. We added their names in Latin, Ancient Greek, English, French, and Italian.

Our goal – and this is important – was to push cleaned data to Wikidata, and then pull it back into our platform – once structured, normalized, and enriched. But as we engaged with Wikidata, we quickly realized that it wasn’t just a passive repository. The community was highly responsive. Feedback came fast – sometimes technical, always helpful – and it forced us to reflect critically on our editorial assumptions.

Let me share two quick examples that illustrate the kind of decisions we had to make.

### **Diodoros (ambiguous entries) @Math [1min]**

Our database initially listed four variants of the name Diodoros, with some entries referring to historical figures that could or could not be the same (“Diodorus, Diodoros”, “Diodore de Tarse, Diodoros le Grammairien”, “Diodoros Zonas de Sardes”, and a lowercase entry, “diodorus”). The Belles Lettres edition mentions at least three poets named Diodoros — and in some cases, it’s impossible to tell which one is being referenced.

Diodoros. Trois poètes de la Couronne de Philippe ont porté ce nom : Diodore Zonas, de Sardes, orateur célèbre du temps de la guerre de Mithridate ; un autre Diodoros de Sardes ; enfin, Diodoros de Tarse. On ne sait auquel attribuer les épigrammes où le gentilicium n’est pas spécifié. (Waltz 2003, II:142–43)

So we decided to keep separate entities when possible and created a generic “Diodoros Epigrammaticus” entry (in line with Perseus conventions) for unattributed cases. This solution may not be perfect, but it’s honest — and it lets uncertainty remain visible rather than artificially resolved.

### **Dionysos (uncertain attribution for common names) @Math [45sec]**

The same logic applied to Dionysios, illustrating the difficulties posed by common names in the manuscript tradition. Many epigrams are attributed simply to “Dionysios”, with no further context – we don’t even know how many distinct authors are represented. Instead of choosing arbitrarily, we retained specific identities when clear (Dionysius of Andros, Dionysius of Cyzicus, Dionysius of Rhodes and Dionysius the Sophist), and created a catch-all “Dionysios” entity where attribution was unclear. Again, the idea was to represent uncertainty, not erase it.

This raises the question of “how do we make ambiguity and uncertainty readable, especially in a digital context?”

## **Gain for our platform and for Wikidata @Math [45sec]**

What started as metadata cleaning quickly became something deeper. We improved many attributions, added inline commentary for uncertain cases (see AG X.38), created new Wikidata items, and contributed multilingual labels.

But more importantly, we encountered a series of modeling challenges that revealed the tensions between interpretive scholarship and structured ontologies.

A good example is the case of pseudo-authors. One of our entries had been labeled in Ancient Greek with the prefix  $\Psi$  - — following modern editorial conventions. But the Wikidata community flagged it as problematic: the prefix is rare in ancient sources and may reflect anachronistic assumptions. The label was removed — and the debate forced us to consider how editorial categories translate (or don't) into semantic web models.

## **Other keywords and future work @Math [1min15]**

Beyond authors, we worked with other types of keywords. For places, we imported data directly from Wikidata. While the coordinates were very helpful, the multilingual labels were less consistent – but we intend to add the information on Wikidata in the following months.

Eventually, we still have to reconcile the other thematic keywords. Some of them are easy, some are not. These categories function as reading tools, but they don't necessarily align with Wikidata's data model. Other tags, like bad breath or personification of weapons, describe literary motifs — relevant for interpretation, but difficult to represent in a structured ontology.

In short, aligning our data with Wikidata is a major step toward openness and interoperability. But reconciling this data reveals just how difficult it is to map scholarly knowledge into structured ontologies — especially when the two systems weren't designed to work together in the first place.

## **Conclusion @Max [1min45]**

To conclude: Wikidata is not just a technical backend. It is a collaborative epistemic space — a platform where academic knowledge is not only stored, but also negotiated, revised, and legitimized.

In the *Anthologia Graeca* project, linking our editorial work to Wikidata meant delegating part of our curatorial authority to a broader community — not as a loss of academic rigor, but as a deliberate move toward a more distributed model of expertise. A model that values scholarly precision, while also embracing the strength of collective intelligence to support high-quality, multilingual, and interoperable data.

This shift invites us to rethink our roles as researchers: not only as producers of content, but also as participants in open knowledge systems — where validation is collaborative, multilingual, and always in progress.

It also raises essential questions: Can academic standards and open platforms like Wikidata truly align? How do we balance transparency with control, or openness with critical nuance?

Rather than smoothing out these tensions, working with Wikidata makes them visible — and in doing so, creates a space for reflection on how knowledge is authored, validated, and shared.

Engaging with such platforms means accepting that authority in the digital age is no longer something we hold — through status, institutional affiliation, or academic citations — but something we negotiate: collectively, contextually, and in dialogue with others.

Thank you.